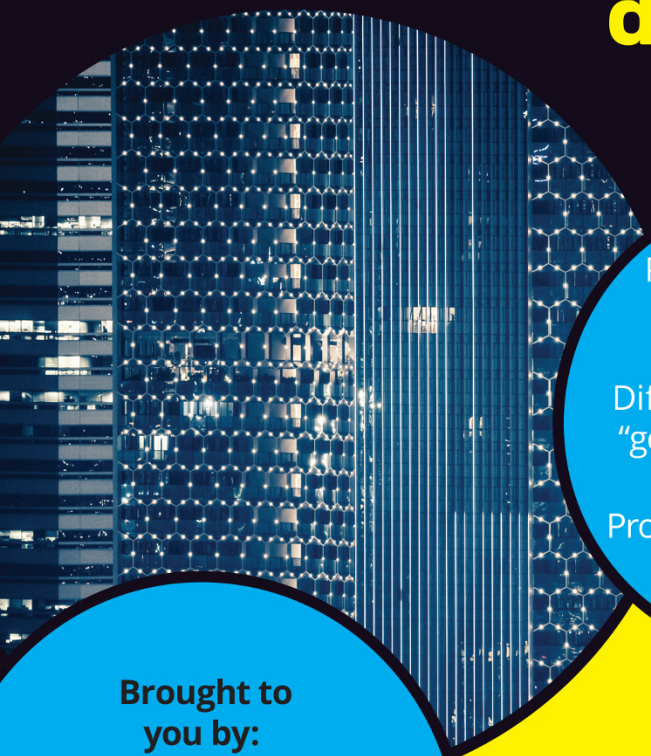
LEARNING MADE EASY

Palo Alto Networks
Limited Edition

Next-Generation Firewalls

for dummies®
A Wiley Brand



Prevent successful cyber attacks

Differentiate between "good" and "bad" apps

Protect apps, users, and data consistently

Brought to you by:

paloalto® NETWORKS

Lawrence Miller

About Palo Alto Networks

We are the global cybersecurity leader, known for always challenging the security status quo. Our mission is to protect our way of life in the digital age by preventing successful cyberattacks. This has given us the privilege of safely enabling tens of thousands of organizations and their customers. Our pioneering Security Operating Platform safeguards your digital transformation with continuous innovation that combines the latest breakthroughs in security, automation, and analytics. By delivering you a true platform and empowering a growing ecosystem of change-makers like us, we provide you highly effective and innovative cybersecurity across clouds, networks, and mobile devices. Across the world, customers love our security and consistently award us the highest loyalty ratings and net promoter scores in the industry. **https://www.paloaltonetworks.com**

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Next-Generation Firewalls

Palo Alto Networks Limited Edition

by Lawrence Miller, CISSP

for dummies®
A Wiley Brand

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Next-Generation Firewalls For Dummies®, Palo Alto Networks Limited Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2019 by John Wiley & Sons, Inc., Hoboken, New Jersey

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permissions.

Trademarks: Wiley, For Dummies, the Dummies Man logo, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN 978-1-119-62401-1 (pbk); ISBN 978-1-119-62402-8 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Project Editor: E. N. Kuball
Acquisitions Editor: Ashley Coffey
Editorial Manager: Rev Mengle
Business Development Representative: Karen Hattan
Production Editor: Siddique Shaik

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Introduction

The rapid evolution of applications, IT infrastructure, and the threat landscape has resulted in a loss of visibility and control for organizations. Devices are proliferating and the network perimeter has all but disappeared, leaving enterprise security teams struggling to safely enable and protect their businesses, customers, and users. With new threats growing in number and sophistication, organizations are finding that traditional security products and approaches are less and less capable of protecting their networks against today's advanced attacks.

At the same time, application development and IT operations teams are accelerating the delivery of new applications to drive business growth by adopting DevOps tools and methodologies, cloud and container technologies, big data analytics, and automation and orchestration. Meanwhile, applications are increasingly accessible. The result is an incredibly complex network that introduces significant business risk. Organizations must minimize this risk without slowing down the business.

A different approach to network security is needed. Defenders need to replace siloed point products with security innovations that are tightly integrated. The cornerstone of this approach is the next-generation firewall (NGFW)!

About This Book

Next-Generation Firewalls For Dummies consists of six chapters that explore the complexity in modern network security (Chapter 1), the evolving threat landscape (Chapter 2), the shortcomings of traditional firewalls and other network security solutions (Chapter 3), the advanced capabilities found in NGFWs (Chapter 4), how to deploy NGFWs (Chapter 5), and what to look for in an NGFW platform (Chapter 6).

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Foolish Assumptions

It's been said that most assumptions have outlived their uselessness, but I assume a few things nonetheless!

Mainly, I assume that you know a few things about network security. Perhaps you're a chief information officer (CIO), chief technology officer (CTO), chief information security officer (CISO), network architect, or security practitioner. Basically, you're some sort of IT, network, or security professional! As such, this book is written primarily for technical readers.

Icons Used in This Book

Throughout this book, I occasionally use special icons to call attention to important information. Here's what to expect:

REMEMBER

The Remember icon points out important information you should commit to your nonvolatile memory, your gray matter, or your noggin!

TECHNICAL STUFF

If you seek to attain the seventh level of NERD-vana, perk up! The Technical Stuff icon explains the jargon beneath the jargon!

TIP

Tips are appreciated, never expected — and I sure hope you'll appreciate these useful nuggets of information.

WARNING

These alerts point out the stuff your mother warned you about (well, probably not), but they do offer practical advice to help you avoid potentially costly or frustrating mistakes.

Beyond the Book

There's only so much I can cover in 80 short pages, so if you find yourself at the end of this book, thinking, "Gosh, this was an amazing book! Where can I learn more?," just go to `www.paloaltonetworks.com`.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Understanding how security became so complex
- Addressing data compromise
- Recognizing the need for speed and simplicity
- Adopting a "never trust, always verify" security posture

Chapter 1
Security Needs to Be Simple

Just as antivirus software has been a cornerstone of PC security since the early days of the Internet, firewalls have been the cornerstone of network security. But today's application and threat landscape renders traditional port-based firewalls largely ineffective at protecting corporate networks and information. Applications are the conduit through which everything flows — a vector for our business and personal lives — along with their associated benefits and risks. Such risks include new and emerging threats, data compromise, and noncompliance.

This chapter explains how traditional firewall technology works, why products based on this legacy approach can't effectively address today's application and threat landscape, why data compromise is so challenging, how complex security processes and technologies increase security risk, and why organizations need to adopt a Zero Trust approach to network security.

Why Is Security So Complicated?

The rapid pace of business today is enabled by modern innovations and trends that include virtualization and containerization, cloud computing, bring your own device (BYOD) mobility,

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

artificial intelligence (AI) and machine learning, and big data analytics, among others. With applications and data now hosted in on-premises data centers and public and private cloud environments, and accessed from anywhere, at any time, on any device, the network perimeter has all but disappeared. As a result, enterprise security teams are struggling to keep pace while safely enabling the business and its users.

A firewall, at its most basic level, controls traffic flow between network segments — at the (disappearing) network perimeter. Many currently deployed firewalls are still port-based firewalls, or some variation (such as stateful inspection) of this basic type of firewall. In the rapid pace of the Internet Age, two decades means the basic technology behind port-based firewalls is medieval. In fact, network security is often likened to the Dark Ages — a network perimeter is analogous to the walls of a castle, with a firewall controlling access — like a drawbridge. And like a drawbridge that is either up or down, a port-based firewall is often limited to just two options for controlling network traffic: allow or block.

Many IT organizations have tried to compensate for deficiencies in traditional port-based firewalls by surrounding them with proxies, intrusion prevention systems (IPSs), URL filtering, and other security point products. This siloed approach has been largely ineffective in today's application and threat landscape and adds needless cost and complexity. Further compounding these challenges, complexity creates additional risks to the enterprise, such as improper configuration of firewalls and other security products. According to a leading analyst firm, 99 percent of firewall breaches will be caused by misconfiguration.

TECHNICAL STUFF

Port-based firewalls (and their variants) use source/destination IP addresses and Transmission Control Protocol (TCP)/User Datagram Protocol (UDP) port information to determine whether a packet should be allowed to pass between networks or network segments. The firewall inspects the first few bytes of the TCP or UDP header in an IP packet to determine the application protocol — for example, Simple Mail Transfer Protocol (SMTP; port 25) and Hypertext Transfer Protocol (HTTP; port 80).

Most firewalls are configured to allow all traffic originating from the trusted network to pass through to the untrusted network, unless it's explicitly blocked by a rule. For example, the Simple Network Management Protocol (SNMP) might be explicitly blocked to prevent certain network information from being

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

inadvertently transmitted to the Internet. This would be accomplished by blocking UDP ports 161 and 162, regardless of the source or destination IP address.

Static port control is relatively easy. Stateful inspection firewalls address *dynamic* applications that use more than one well-defined port (such as File Transfer Protocol [FTP] ports 20 and 21). When a computer or server on the trusted network originates a session with a computer or server on the untrusted network, a connection is established. On stateful packet inspection firewalls, a dynamic rule is temporarily created to allow responses or replies from the computer or server on the untrusted network. Otherwise, return traffic needs to be explicitly permitted, or access rules need to be manually created on the firewall (which usually isn't practical).

All of this works well as long as everyone plays by the rules. Unfortunately, the rules are more like guidelines, and not everyone using the Internet is nice! The Internet has spawned a new generation of applications being accessed by everyone for both personal and business use. Many of these applications help improve user and business productivity but increase security and business risk — for example, data leaks and compliance. And many of these applications incorporate techniques, such as using nonstandard ports, port hopping, and tunneling, to evade traditional port-based firewalls.

To address these challenges, organizations need to replace complex and risky siloed point security solutions built on port-based technology and provide consistent security everywhere: at headquarters, at branch offices, at mobile and remote locations, and in the data center and cloud.

Data Compromise Is a Problem

Large-scale, public exposures of sensitive or private data are far too common. Numerous examples of accidental and deliberate data loss continue to regularly make nightmare headlines, exposing the compromise of millions of credit card numbers by major retailers, Social Security numbers leaked by government agencies, protected health information (PHI) disclosed by health-care organizations, and other sensitive information lost by employers in practically every industry. Unfortunately, such incidents are not isolated. In many of these cases, sensitive data was compromised

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

starting with an application that was expressly prohibited by policy but not adequately enforced with technology, or via an application that was allowed, but also carried a threat that gained a foothold by automatically infecting a computer or fooling a user. Other risks to data include data sabotage (or destruction) and the use of ransomware that encrypts important data, rendering it unusable, unless a hefty ransom is paid (CryptoLocker and WannaCry are two examples).

With respect to data loss, data loss prevention (DLP) is sometimes held up as a solution. Unfortunately, given the scope, size, and distributed nature of most organizations' data sets, just discovering where sensitive data is and who owns it is an insurmountable challenge. Adding to this challenge, questions regarding access control, reporting, data classification, data at-rest versus data in-transit, desktop and server agents, and encryption abound. As a result, many DLP initiatives within organizations progress slowly and eventually falter. And DLP does nothing to prevent data sabotage (because it was never designed to address this problem).

Controlling the applications that are used to compromise data, either directly or as part of a larger "chain of events" is foundational to securing organizations. Exerting that control at trust boundaries is ideal — whether the demarcation point is between

- The enterprise network and the Internet (the traditional "network perimeter")
- Network segments within the corporate local area network (LAN)
- Public and private cloud gateways
- Internal users and resources in the data center
- Individual systems and applications within different trust zones in the data center
- Different virtual machines (VMs) on the same physical host (micro-segmentation)

Micro-segmentation enables an enterprise to logically divide network and data center resources at a highly granular level (even down to individual VM workloads) and define security policies and controls that are dynamically applied to those resources, regardless of their location at any given time (such as physical host, data center, private cloud, or public cloud).

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

The firewall sits in the perfect location, seeing all traffic traversing different networks and network segments. Unfortunately, legacy port- and protocol-based firewalls can't do anything about any of this — being ignorant of applications, users, and content.

To effectively address data compromise with a firewall solution, organizations should

- Gain control over the applications on their network — thus, limiting the avenues of data loss or compromise
- Scan the applications they want on their networks, for sensitive or private data, or to detect behaviors in a multi-stage attack designed to steal or sabotage data
- Understand which users are initiating application transactions and why
- Implement appropriate control policies and technology to prevent accidental or intentional data loss or compromise

If organizations could control applications and the flow of sensitive or private data in the network, many of the data loss incidents that regularly make the news could be prevented. Unfortunately, legacy security infrastructures, with traditional port-based firewalling as their basis, are ill equipped to provide this functionality.

Go Fast and Manage Risk

Modern businesses need speed and agility to compete effectively. As the pace of business accelerates, you also need to manage risk — ideally maintaining or reducing your current level of risk. Going fast while effectively managing risk requires you to change the equation by simplifying your processes and technology. Complexity not only slows you down and increases uncertainty, but also adds security risk.

When a process is complicated, there are more opportunities for things to go wrong. Simplifying processes by removing extra steps and eliminating manual tasks speeds execution.

Likewise, deploying multiple security and firewall technologies doesn't improve security; instead, it adds risk due to the management and operational complexity. "Defense in depth" has become "defense ad nauseum," with unintended and serious

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

consequences for the organization. The situation is similar in the public cloud, with most breaches today being cause by exposed keys, weak passwords, and inappropriate access configuration.

Simplicity requires three things:

- **Consistent visibility and control:** You must know exactly what's happening on your network. You need to inspect Secure Sockets Layer (SSL) and Transport Layer Security (TLS) encrypted network traffic and Domain Name System (DNS) tunneling traffic. You need visibility into mobile user traffic and inside Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS) applications and solutions. But it's not enough to just have visibility — you also need control, so you can take action based on what you see happening on your network.
- **Tightly integrated tools that replace manual effort:** You need to replace disconnected and siloed tools that require manual effort to configure and operate together. How many security tools do you have in your organization? In many organizations, 10 to 20 is common; larger organizations may use 40 to 50. Many vendors offer "platforms" to combine these disconnected tools, but calling something a platform is not enough. The integrations must be built in to deliver the efficiencies that reduce complexity and increase response speed.
- **Analytics that drive automatic action:** Analytics is important to every area of the business (for example, to understand how to better service your customers and improve your supply chain). Analytics is particularly useful in dealing with the volume and sophistication of modern threats. Attackers use automation and analytics. It's time to do the same to accelerate and optimize prevention and response.

Automating manual processes with the right technology enables you to accelerate execution and reduce — or even eliminate — human error, thereby reducing overall risk to the organization.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Trust Makes You Vulnerable

In the not-too-distant past, security vendors, network architects, and security practitioners described networks in terms of the "untrusted" public Internet and the "trusted" internal corporate network with firewalls deployed at the network perimeter. But as discussed earlier in this chapter, the network perimeter has become a relic of a bygone era when everything was simple: black or white, good or bad, "trusted" or "untrusted."

The reality is that attackers have always exploited relatively weak security designs that relied on the firewall as the arbiter of trust between the Internet and the corporate network. Once inside, attackers had — and continue to have — free rein in the data center and on the network, because trust is assumed. This threat is further exacerbated by the fact that traditional firewalls deployed at the network perimeter only inspect *north–south traffic* (that is, traffic passing between the private corporate network and the public Internet). These firewalls have no visibility into *east–west traffic* (traffic between systems and applications inside the data center), which today constitutes the majority of data center network traffic.

Forrester Research created the Zero Trust model for network security which implements the security principle of "least privilege" by adopting a default "never trust, always verify" approach to network security. You can find out more about Zero Trust in Chapter 5.

TIP

Palo Alto Networks is ranked as a leader in the fourth-quarter 2018 report, *The Forrester Wave: Zero Trust eXtended (ZTX) Ecosystem Providers.*

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Identifying applications as good, bad, or good and bad
- Understanding accessibility tactics
- Recognizing the speed and sophistication of today's threats

Chapter 2
Defining the Threat Landscape

Network security used to be relatively simple — everything was more or less black and white — either clearly bad or clearly good. Business applications constituted good traffic that should be allowed, while pretty much everything else constituted bad traffic that should be blocked.

Problems with this approach today include the fact that applications have become

- Increasingly "gray" — classifying types of applications as good or bad is not a straightforward exercise.
- More difficult to accurately identify based on traditional port and protocol assignments — threat adversaries take advantage of this difficulty to attack organizations that rely on port-based firewall rules.
- The attack vector of choice for today's cybercriminals and threat developers who use applications as unwitting carriers of malicious payloads.

This chapter explores the evolving threat landscape, the blurring distinction between user and business applications, and the

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

strategic nature of many of these applications (and their associated risks) for businesses today.

Applications Are Not All Good or All Bad

The application landscape is constantly evolving and has changed dramatically for organizations over the years. Corporate productivity applications have been joined by a plethora of personal and consumer-oriented applications that are often available as Software as a Service (SaaS) or web-based applications. This convergence of corporate infrastructures and personal technologies is being driven by two popular and important trends — *consumerization* and *bring your own device* (BYOD).

The process of consumerization occurs as users increasingly find personal technology and applications that are more powerful or capable, more convenient, less expensive, quicker to install, and easier to use than corporate IT solutions. These user-centric "lifestyle" applications and technologies enable individuals to improve their personal efficiency, handle their nonwork affairs, and maintain online personas, among other things.

Catering to this demand, technology vendors and developers enjoy vast economies of scale and the pervasive benefits of viral marketing.

Closely related to consumerization is BYOD — organizations permit their employees to use their own personal devices, primarily smartphones and tablets, for work-related purposes. More often than not, the same applications used for social interaction on personal devices are being used for work-related purposes. And as the boundary between work and their personal lives becomes less distinct — particularly for millennial workers — users are practically demanding that these same tools be available to them in their workplaces.

The rapid adoption of many popular SaaS and mobile applications is often driven by users, not by IT. The ease with which they can be accessed, combined with the fact that today's knowledge workers are accustomed to using them, points toward a continuation of the consumerization trend and a growing "shadow IT" culture in which individuals and departments use both sanctioned and unsanctioned applications.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

The applications driven by consumerization combine with those supported by IT, resulting in a wide variety of application types in organizations today. Examples of these applications include

- Collaboration and cloud storage tools such as Box, Dropbox, Google Docs, iCloud, Microsoft Office 365, and OneDrive
- Web-based email such as Gmail, Outlook.com, and Yahoo! Mail
- Content management tools such as SharePoint
- Customer relationship management (CRM) portals such as Salesforce and SugarCRM
- Social networks such as Facebook and LinkedIn
- Web publishing tools such as YouTube
- Unified messaging tools such as Skype and Vidyo
- Posting tools such as Twitter
- Anonymizers and proxies such as Tor and UltraSurf
- Remote access tools such as Ammyy, LogMeIn, Remote Desktop Protocol (RDP), and TeamViewer

WARNING

The use of anonymizers and proxies on any network should be considered risky and suspect.

WARNING

Remote access tools can be both good and bad. They're valuable productivity tools for IT administrators and support technicians, but also prone to exploit by attackers in order to control systems.

Unsure of how to leverage these trends in their business processes, many organizations either implicitly allow these SaaS-based and mobile applications simply by ignoring their use in the workplace, or explicitly prohibit their use, but are then unable to effectively enforce such policies with traditional firewalls and security technologies. Neither of these two approaches is ideal, and both incur inherent risks for the organization. In addition to lost productivity, adverse issues for the organization include

- Creating a "shadow IT" subculture of back-channel or underground workflow processes that are critical to the business's operations, but are known only to a few users and fully dependent on personal technologies and applications

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- Introducing new risks to the entire networking and computing infrastructure, due to the presence of unknown, and, therefore, unaddressed and unpatched, vulnerabilities, as well as threats that target normal application and user behavior — whether a vulnerability exists in the application or not
- Being exposed to noncompliance penalties for organizations that are subject to increasingly complex and stringent regulatory requirements (see Chapter 1 for several examples)
- Having employees circumvent controls with external proxies, encrypted tunnels, and remote desktop applications, making it difficult, if not impossible, for security and risk managers to see the risks they're trying to manage

The challenge is not only the growing diversity of the applications, but also the inability to clearly and consistently classify them as good or bad. Although many are clearly good (low risk, high reward), and others are clearly bad (high risk, low reward), most are somewhere in between. Moreover, the end of the spectrum that these applications fall on can vary from one scenario to the next and from user to user or from session to session.

For example, using a social networking application to share product documentation with a prospective customer would be "good" (medium risk, high reward), while using the same application to forward details of an upcoming release to a "friends list" that includes employees of a competitor would be "not so good" (high risk, no reward).

Indeed, many organizations now use a variety of social networking applications to support a wide range of legitimate business functions, such as recruiting, research and development, marketing, and customer support — and many are even inclined to allow the use of lifestyle applications, to some extent, as a way to provide an "employee friendly" work environment and improve morale.

Many companies are also seeing significant benefits from the use of these applications and technologies, including an increased ability to share ideas, more rapid access to knowledge experts, and a reduction in travel, operations, and communications costs. Today's network security solutions, therefore, must be able not only to distinguish one type of application from the next, but also to account for other context surrounding its use (for example, file transfers and URL access) and to vary the resulting action that will be taken accordingly.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

ENABLING FACEBOOK USAGE WHILE PROTECTING THE BUSINESS

Despite recent controversy surrounding its privacy policies, Facebook continues to expand its influence from the personal world to the corporate world, as employees use these applications to get their jobs done. At the same time, many organizations are looking at the more than 2.4 billion Facebook users as an opportunity to conduct research, execute targeted marketing, gather product feedback, and increase awareness. The end result is that Facebook (and other social networks, such as LinkedIn) can help organizations improve their bottom line.

However, formally enabling the use of Facebook introduces several challenges to organizations. Many organizations are unaware of how heavily Facebook is being used, or for what purpose. In most cases, policies governing specific usage are nonexistent or unenforceable. Finally, users tend to be too trusting, operating in a "click now, think later" mentality, which introduces significant security risks.

Like any application that is brought into the enterprise by end users, blindly allowing Facebook may result in propagation of threats, loss of data, and damage to the corporate reputation. Blindly blocking Facebook is also an inappropriate response because it may play an important role in the business and may force users to find alternative means of accessing it (such as proxies, circumvention tools, and others).

Organizations should follow a systematic process to develop, enable, and enforce appropriate Facebook usage policies while simultaneously protecting network resources:

- **Find out who's using Facebook.** A "corporate" Facebook presence may already be established by marketing or sales, so it's critical that IT determine which social networking applications are in use, who is using them, and the associated business objectives.
- **Develop a corporate Facebook policy.** When Facebook usage patterns are determined, organizations should engage in discussions regarding what should and should not be posted. Educating users on the security risks associated with Facebook is another important element to encouraging usage for business purposes.
- **Use technology to monitor and enforce policy.** The outcome of each of these discussions should be documented with an explanation of how IT will apply security policies to safely and securely enable use of Facebook within enterprise environments.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Applications: "I'm Not a Number!"

Although "distinguishing one type of application from the next" sounds simple, it really isn't — for a number of reasons. In order to maximize their accessibility and use, many applications are designed from the outset to use standard ports, such as Transmission Control Protocol (TCP) ports 80 (Hypertext Transfer Protocol [HTTP]) and 443 (Hypertext Transfer Protocol Secure [HTTPS]), that are commonly allowed through legacy port-based firewalls that see applications as little more than a number. You've heard the expression, "If all you have is a hammer, everything looks like a nail," right? Well, to a port-based firewall, increasingly every application looks like HTTP or HTTPS! And because anything can run over these ports, your firewall provides you no visibility.

Other applications use a variety of techniques in an attempt to run anywhere, at any time. Common techniques include the following:

- **Port hopping,** where ports/protocols are randomly shifted over the course of a session
- **Use of nonstandard ports,** such as running Yahoo! Messenger over TCP port 80 (HTTP) instead of the standard TCP port for Yahoo! Messenger (5050)
- **Tunneling within commonly used services,** such as when sharing files or using messaging applications like Telegram Messenger
- **Hiding within Secure Sockets Layer (SSL) encryption,** which masks the application traffic, for example, over TCP port 443 (HTTPS)

WARNING

These techniques are also used by attackers for malicious purposes to evade detection by port-based firewalls.

At the same time, enterprise users are increasingly embracing SaaS and web-based applications and services such as Salesforce, WebEx, and Google Apps — which often initiate in a browser but then switch to more client/server-like behavior (rich client, proprietary transactions, and others).

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

The result of the shift to SaaS and web-based applications is that HTTP and HTTPS now account for approximately two-thirds of all organizational traffic. This is not a problem, per se, but it does exacerbate an inherent weakness of traditional security infrastructure. Specifically, the wide variety of higher-order applications riding on top of HTTP and HTTPS — whether or not they actually serve a legitimate business purpose — are practically indistinguishable for older port-based firewalls. The negative impact of organizations further losing control over their network communications is clear and underlines the fact that the application landscape has evolved dramatically.

CLOUD-BASED SAAS APPLICATIONS: I CAN'T SEE CLEARLY NOW

Organizations are adopting SaaS-based application services at a breakneck pace. These applications continue to redefine the network perimeter, providing critical functionality and increasing productivity, but at the same time introducing potential new security and data risks if not properly controlled.

In most organizations that use SaaS applications, users are provided access to a specific list of services that the organization has deemed acceptable or suitable for business purposes. However, given the large number of unique SaaS applications that are readily available on the Internet, it's likely that many users aren't strictly complying with such usage policies and are instead using unsanctioned SaaS applications at work. This further increases the risk of data leakage to organizations, due to the lack of visibility from regular logs or notifications from unauthorized SaaS applications, as well as additional risk of intermeshing users' personal and work emails, which may create situations where a user's personal email account is attacked and the attacker is then able to steal data or compromise the user's work email account.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Threats Are Coming Along for the Ride

The increasing prevalence of application-layer attacks is yet another disturbing trend. Email and web browsers are still the main attack vectors today, with malicious content either attached or downloaded as an executable or macro-based file. The malicious use of remote access applications is another significant attack vector. Threats that directly target applications can pass right through the majority of enterprise defenses, which have historically been built to provide network-layer protection. Threat developers exploit the same methods (described in the previous section) to infiltrate networks that application developers utilize to promote ease of use and widespread adoption, such as tunneling within applications.

WARNING

Legitimate applications are being used by attackers to spread malware.

The evasion techniques built into these and many other modern applications are being leveraged to provide threats with "free passage" into enterprise networks. So, it's no surprise that more than 80 percent of all new malware and intrusion attempts are exploiting weaknesses in applications, as opposed to weaknesses in networking components and services. Together with the implicit trust that users place in their applications, all these factors combine to create a "perfect storm." The motivation for attackers has also shifted — from gaining notoriety to political activism, espionage, and making money. The name of the game today is information theft. Consequently, it's no longer in an attacker's best interests to devise threats that are "noisy" or that are relatively benign. To be successful, a thief must be fast or stealthy — or both.

For those attackers who favor speed over sophistication — speed of initial threat generation, speed of modification, and speed of propagation — the goal is to develop, launch, and quickly spread new threats immediately on the heels of the disclosure of a new vulnerability. The resulting zero-day and near-zero-day exploits then have an increased likelihood of success because reactive countermeasures, such as patching and those tools that rely on threat signatures (such as antivirus software and intrusion prevention), have trouble keeping up — at least during the early phases of a new attack.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

THE (RE)RISE OF MACRO MALWARE

In 1995, the first macro-based malware, WM/Concept, was unleashed upon the public and began the initial wave of macro-based malware targeting Microsoft Word and Excel. Twenty years later, cybercriminals have rediscovered macro-based malware and are once again using it as another tool in their arsenal.

What is a macro?

Macros were originally developed for the Microsoft Office suite as a way to automate repetitive tasks or share tasks between different users. The system was designed so that a user could use a simple feature to record a repetitive task, which would then automatically be transcribed into Visual Basic for Applications (VBA). The macro automated the task, and the VBA code could then be shared with other users.

Unfortunately, macros were not designed with security in mind; functionality was the main goal, and macros allowed users to be more productive by speeding up repetitive tasks. Although the intentions of macros in Microsoft Office documents were altruistic, the unfortunate side effect was the creation of an easy-to-use and effective vehicle for malicious code.

The most famous and well-known macro-based malware was the Melissa virus in 1999. It was distributed within a Word document that would gather the first 50 entries from a user's address book and then mail a copy of the macro-infected Word document to each entry via Microsoft Outlook. When the recipients opened the document, the cycle would continue ad nauseam. Due to the overwhelming number of infected systems attempting to send out emails, the Melissa virus placed many major email servers into a denial-of-service state.

Where are we now?

In response to the Melissa virus and other macro malware, Microsoft put multiple mitigations in place to prevent the spread of macro-based malware. In Office 2003, only digitally signed macros could be run by default. In Office 2007, the letter *m* was appended to the usual

(continued)

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

(continued)

Office file extensions (`.docxm`, `.xlsxm`, `.pptxm`) to signify that the file contained a macro. Finally, in Office 2013, macros were simply turned off by default, showing users a notification if a macro was embedded in the document they had opened. The actions taken by Microsoft significantly reduced macro-based malware infections and, in turn, reduced the popularity of macro-based malware usage by cybercriminals.

No good deed goes unpunished, however. In the last decade, a new generation of users who have never used macros or are even aware of what they are due to the dormancy of macros in general has emerged. Users have a tendency to have a singular goal in mind, which is to accomplish the given task at hand. This causes them to ignore warnings or pop-up messages indicating potential danger because, to them, these buttons and dialog boxes are simple barriers to their productivity. The lack of awareness and a focus on getting to the user's desired content or task has led to a sudden resurgence in the usage of macro-based malware as users unwittingly enable macros in Office documents more and more often.

This speed-based approach is facilitated in large part by the widespread availability of threat development websites, toolkits, and frameworks. Unfortunately, another by-product of these resources is the ability to easily and rapidly convert "known" threats into "unknown" threats — at least from the perspective of signature-based countermeasures. This transformation can be accomplished either by making a minor tweak to the code of a threat, or by adding entirely new propagation and exploit mechanisms, thereby creating what is commonly referred to as a *blended threat.*

Many of today's threats are built to run covertly on networks and systems, quietly collecting sensitive or personal data, and going undetected for as long as possible. This approach helps to preserve the value of the stolen data and enables repeated use of the same exploits and attack vectors. As a result, threats have become increasingly sophisticated. Rootkits, for example, have become

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

more prevalent. These kernel-level exploits effectively mask the presence of other types of malware, enabling them to persistently pursue the nefarious tasks they were designed to accomplish (such as intercepting keystrokes).

Encryption is increasingly used to secure not just sensitive or private information, but practically all traffic traversing enterprise networks. However, organizations are essentially left blind to any security threats contained inside encrypted traffic. Attackers exploit this lack of visibility and identification to hide within encrypted traffic and spread malware. Even legitimate websites that use SSL can be infected with malware. Moreover, attackers increasingly use SaaS applications to deliver malware. For example, an attacker can place a malicious file on a website with encryption and host a file to be downloaded.

Examples of malware transferred over encrypted traffic include:

- **Upatre:** Steals credentials
- **Dridex:** Transfers funds illegally
- **Ehdoor:** Steals sensitive information

WARNING

Without the ability to decrypt, classify, control, and scan SSL-encrypted traffic, it's impossible for an organization to adequately protect its business and its valuable data from modern threats.

Threats to enterprise networks and computing environments include:

- Ransomware
- Credential theft
- Domain Name System (DNS)–based attacks
- Targeted attacks and APTs

Ransomware

Ransomware has existed in various forms for decades, but in the last few years, criminals have perfected the key components for

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

these types of attacks. Ransomware uses malware to encrypt a victim's data until a ransom is paid — usually in cryptocurrency. Ransomware has become a multimillion-dollar criminal business targeting both individuals and corporations. Due to its low barriers to entry and effectiveness in extorting ransom payments from its victims, the spread of ransomware has increased exponentially in recent years. A typical ransomware attack consists of the following steps:

1. **Compromise and control a system or device.**

   Most ransomware attacks begin by using social engineering to trick users into opening an attachment or viewing a malicious link in their web browser. This allows attackers to install malware onto a system and take control.

2. **Prevent access to the system.**

   Attackers will either identify and encrypt certain file types or deny access to the entire system.

3. **Notify the victim.**

   Though seemingly obvious, attackers and victims often speak different languages and have varying levels of technical capabilities. Attackers must alert the victim about the compromise, state the demanded ransom amount, and explain the steps for regaining access.

4. **Accept ransom payment.**

   To receive payment while evading law enforcement, attackers utilize cryptocurrencies such as Bitcoin for the transaction.

5. **Restore full access (usually).**

   Attackers must return access to the device(s). Failure to restore the compromised system(s) destroys the effectiveness of the scheme — no one would be willing to pay a ransom if he didn't believe access to his data would be restored.

Credential theft

According to Forrester Research, at least 80 percent of data breaches today involve compromised privileged credentials. Credential theft has become so prevalent in the attackers' playbook that it's often said that attackers no longer hack into a

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

target network — they simply log in. The primary techniques that attackers use to steal credentials include

- Social engineering
- Credential phishing and spam
- Reusing stolen passwords or shared credentials
- Brute force
- Security question reuse

RANSOMWARE: LOCKERGOGA

The LockerGoga ransomware was first publicly reported in January 2019 by Bleeping Computer, which tied the malware to an attack against French engineering company Altran Technologies. Several variants have since been found in the wild, where they were used in attacks against Norwegian aluminum manufacturer Norsk Hydro and two chemical companies: Hexion and Momentive.

Currently, LockerGoga does not support any worm-like capabilities that would allow it to self-propagate by infecting additional hosts on a target network. LockerGoga has been observed moving around a network via the Server Message Block (SMB) protocol, which indicates the actors simply manually copy files from computer to computer.

LockerGoga's developers continue to add capabilities and launch new attacks. The addition of WS2_32.dll and use of undocumented Windows application programming interface (API) calls indicates a level of sophistication beyond typical ransomware authors. The former could lead to the eventual inclusion of command and control (C2) communication or automated propagation, and the latter requires some working knowledge of Windows internals.

These features raise more questions about the actor's intent, as ransomware is typically one of the least advanced forms of malware. Are they motivated by profits or something else? Has the motive changed over time? Why would developers put so much effort into their work only to partially encrypt files? Why do they include an email address, rather than seeking payment through more frequently used cryptocurrencies?

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

CREDENTIAL THEFT: SHAMOON 2

Palo Alto Networks Unit 42 researchers have been following the Shamoon 2 attacks closely since November 2016. Credential theft is a key part of Shamoon 2 attacks. Shamoon 2 enters and spreads through an organization in three stages:

1. Shamoon 2 attackers access and compromise a single system in the network using Remote Desktop Protocol (RDP) with stolen, legitimate credentials. This becomes their distribution server: They download their tools and malware to this system.
2. Attackers execute commands on the distribution server to connect to specific, named systems on the network using the stolen, legitimate credentials, and infect them with the Disttrack malware.
3. The Disttrack malware will execute on those named systems the attacker has successfully infected. The Disttrack malware will attempt to connect to and spread itself to up to 256 IP addresses on its local network. Any systems successfully infected in this stage will also attempt to infect up to 256 IP addresses on their local networks.

Shamoon 2 attacks are very targeted to a specific region, but it would be a mistake to disregard the threat that Shamoon 2 demonstrates. Shamoon 2 attackers are using a rudimentary, but effective, distribution system of their own making. The power of their attack doesn't lie in the tools they use, but in their ability to obtain and abuse legitimate credentials.

DNS-based attacks

The DNS is a massive and often overlooked attack surface present in every organization. According to the Palo Alto Networks Unit 42 threat research team, almost 80 percent of malware uses DNS to initiate command-and-control (C2) communications. Unfortunately, security teams often lack basic visibility into how threats use DNS to maintain control of infected devices. Adversaries take advantage of the ubiquitous nature of DNS to abuse it at multiple points of an attack, including reliable C2.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

DNS-BASED ATTACKS: OILRIG

OilRig is an active, organized threat group first discovered by the Palo Alto Networks Unit 42 threat research team. Operating primarily in the Middle East, OilRig carefully targets organizations to further its regional strategic goals across multiple industries, including supply-chain-based attacks. As part of its adversary playbook, the group employs sophisticated, custom DNS tunneling for C2 and data exfiltration. The use of tunneling includes

- **ALMA Communicator Trojan,** which uses DNS tunneling to receive commands from the adversary and exfiltrate data. The malware employs specially crafted subdomains to send data to the C2 server and specific Internet Protocol version 4 (IPv4) addresses to transmit data from the C2 to the Trojan over DNS requests.
- **Helminth PowerShell-based Trojan,** which can obtain files from a C2 server using a series of DNS text (TXT) queries repeated every 50 milliseconds, essentially building malware on victim systems through hard-to-detect increments sent over DNS.

OilRig's use of DNS tunneling allows the group to establish reliable C2 that can potentially evade existing defenses to carry out further stages of the attack.

Security teams struggle to keep up with new malicious domains and enforce consistent protections for millions of emerging domains at once. It's impossible for enterprise network and security teams to keep up with the high volume of malicious domains, let alone advanced tactics like DNS tunneling for stealthy data theft.

Targeted attacks and APTs

Targeted attacks and APTs against specific organizations or individuals are another major concern. In this case, attackers often develop customized attack mechanisms to take advantage of the specific equipment, systems, applications, configurations, and even personnel employed in a specific organization or at a given location, and quietly collect sensitive data over extended periods.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

APTs are a class of threats that often combine advanced malware and botnet components to execute a far more deliberate and potentially devastating attack than other types of attacks. As the name applies, an APT has three defining characteristics:

- **Advanced:** In addition to advanced malware and botnets, the attackers typically have the skills to develop additional exploitation tools and techniques, and may have access to sophisticated electronic surveillance equipment, satellite imagery, and even human intelligence assets.
- **Persistent:** An APT may persist over a period of many years. The attackers pursue specific objectives and use a low-and-slow approach to avoid detection. The attackers are well organized and typically have access to substantial financial backing to fund their activities, such as a nation-state or organized crime.
- **Threat:** An APT is a deliberate and focused, rather than opportunistic, threat that can cause real damage.

REMEMBER

A *botnet* is a broad network of malware-infected endpoints (bots) working together and controlled by an attacker through C2 infrastructure.

The increasing speed and sophistication of threats emphasize the need for proactive countermeasures with extensive visibility and control at the application layer of the network computing stack.

CARBANAK: THE GREAT BANK ROBBERY

Carbanak is one example of a targeted attack that began in August 2013 and is currently still active despite the arrest of the alleged Carbanak "mastermind" in March 2018. The attackers have sent spear-phishing emails with malicious Control Panel File (CPL) attachments or Word documents exploiting known vulnerabilities. When an initial system has been compromised, additional reconnaissance is performed to identify ATMs, financial accounts, or other areas where money can be transferred for eventual extraction. Each raid has lasted two to four months. To date, the attackers have targeted more than 100 financial institutions and businesses, causing aggregated losses estimated at more than $1 billion.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Inspecting weaknesses in legacy port-based firewalls
- Examining the shortcomings of intrusion prevention
- Addressing device sprawl

Chapter 3
Recognizing the Challenges of Legacy Security Infrastructures

As the application and threat landscape has quickly evolved, the impact within many organizations is that security has grown too complex and IT has lost control. The inability of existing security infrastructures to effectively distinguish good or desirable applications from those that are bad or unwanted forces most IT shops to take an inflexible and untenable "all-or-nothing" approach to security, in which they do one of the following:

- **Take a permissive stance.** This approach ensures the accessibility of important applications, but also allows unwanted applications and threats on the corporate network.
- **Just say "no."** This approach maintains a high state of security, but at the risk of limiting business agility and productivity, alienating users and business units, and creating an underground subculture of backdoor processes to circumvent security controls.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Instead, IT needs the capability to exert granular control and provide in-depth protection down to the level of individual applications, files, and files within applications, in order to confidently say "yes" to legitimate requests from the business and its end users. Unfortunately, traditional network security infrastructures have failed to keep pace and are unable to provide this functionality.

In this chapter, you find out how the new application and threat landscape has challenged these legacy security devices, particularly firewalls, beyond their capability to effectively protect today's networks.

What Ever Happened to the Firewall?

Have you noticed that nobody gets excited about a firewall anymore? There was a time when the firewall was the single most important security device in your network. So, what happened?

The answer is a bit of a cliché, but the Internet has changed everything! Many firewalls still being purchased and deployed today are based on technology that predates the Internet and are, thus, not designed to handle modern threats. Years ago, most firewalls did a pretty good job of controlling traffic in and out of corporate networks. That's because application traffic was generally well behaved. Email would typically flow through port 25, File Transfer Protocol (FTP) was assigned to port 20, and the whole "web surfing" was hanging, uhhh, port 80. Everybody played by the rules that "ports + protocols = applications" and the firewall had everything under control. Blocking a port meant blocking an application. Nice and simple.

Unfortunately, the Internet has never really been nice and simple. And that is truer today than ever before. Today, the Internet often accounts for 70 percent or more of the traffic on your corporate network. And it's not just port 80 web surfing. Typically, 20 percent to 30 percent of it is encrypted Secure Sockets Layer (SSL) traffic on port 443. Even worse, a plethora of new Internet applications insist on making their own rules. They wrap themselves in other protocols, sneak through ports that don't belong to them, and bury themselves inside SSL/Transport Layer Security (TLS) tunnels. In short, they just don't play fair.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

All these applications carry some inherent risk to your organization. And they play host to clever new threats that can slip through your firewall undetected. Meanwhile, your firewall just sits there like nothing is wrong because it's still playing by rules that don't exist anymore!

Port-based firewalls have poor vision

Because they're deployed in-line at critical network junctions, firewalls see all traffic and, therefore, are the ideal resource to provide granular access control. The problem, however, is that most firewalls are "farsighted." They can see the general shape of things, but not the finer details of what is actually happening. This is because they operate by inferring the application-layer service that a given stream of traffic is associated with, based on the port number used in the packet's header, and they only look at the first packet in a session to determine the type of traffic being processed, typically to improve performance. They rely on a convention — not a requirement — that a given port corresponds to a given service (for example, Transmission Control Protocol [TCP] port 80 corresponds to Hypertext Transfer Protocol [HTTP]). As such, they're also incapable of distinguishing between different applications that use the same port/service (see Figure 3-1).

The net result is that traditional, "port-based" firewalls have basically gone blind. Besides being unable to account for common evasion techniques such as port hopping, protocol tunneling, and the use of nonstandard ports, these firewalls simply lack the visibility and intelligence to discern which network traffic

- Corresponds to applications that serve a legitimate business purpose
- Corresponds to applications that can serve a legitimate business purpose but, in a given instance, are being used for unsanctioned activities
- Should be blocked because it includes malware or other types of threats, even though it corresponds to legitimate business activities

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

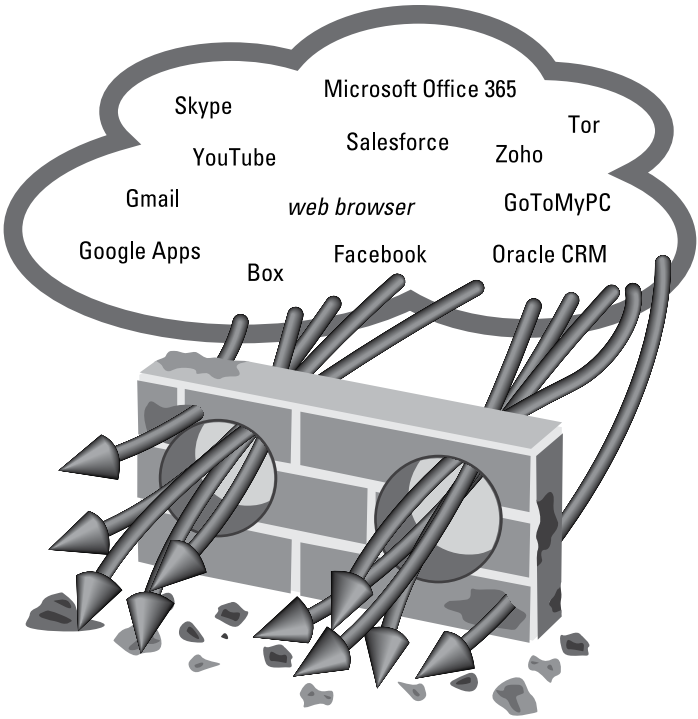


FIGURE 3-1: Port-based firewalls can't see or control applications.

On top of everything else, their control model is typically too coarse-grained. Said firewalls can either block or allow traffic, but offer little variation in between to craft a more appropriate response for all the "gray" applications that enterprises would ultimately like to support — for example, by allowing certain functions or file transfers within an application but not others, allowing but also applying traffic-shaping policies, allowing but scanning for threats or confidential data, or allowing based on users, groups, or time of day.

It doesn't really help matters that the most common steps taken to address the inadequacies of traditional firewalls, for all intents and purposes, have been completely unsuccessful.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Bolt-on functionality is fundamentally flawed

Many purveyors of traditional firewalls have attempted to correct the farsighted nature of their products by incorporating deep packet inspection (DPI) capabilities. On the surface, adding a measure of application-layer visibility and control in this manner appears to be a reasonable approach. However, the boost in security effectiveness that can be achieved in most cases is only incremental because the additional capability is being "bolted on," and the foundation it's being bolted onto is weak to begin with. In other words, the new functionality is added on rather than integrated, and the port-based firewall, with its complete lack of application awareness, is still used for initial classification of all traffic. This leads to problems and limitations, such as the following:

- **Applications that should not be on the network are allowed onto the network.**
- **Not everything that should be inspected necessarily gets inspected.** Because the firewall is unable to accurately classify application traffic, deciding which sessions to pass along to the DPI engine becomes a hit-or-miss proposition.
- **Security posture gets limited.** The bolted-on application classification ability often doesn't get shared with later enforcement capabilities (for example, file transfer control). This makes it impossible for those enforcement options to be precisely applied "per application."
- **Policy management gets convoluted.** Rules on how to handle individual applications essentially get "nested" within the DPI portion of the product — which itself is engaged as part of a higher/outer-level access control policy.
- **Inadequate performance forces compromises to be made.** Inefficient use of system resources and central processing unit (CPU) and memory intensive application-layer functionality can put considerable strain on the underlying platform. To account for this situation, administrators can only implement advanced filtering capabilities selectively.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Firewall "helpers" don't help

Over the years, enterprises have also tried to compensate for their firewalls' deficiencies by implementing a range of supplementary security solutions, often in the form of stand-alone appliances. Intrusion prevention systems (IPSs), antivirus gateways, web filtering products, and application-specific solutions — such as a dedicated platform for instant messaging security — are just a handful of the more popular choices. Unfortunately, the outcome is disappointingly similar to that of the DPI approach, with an additional twist.

Not everything that should get inspected does because these firewall helpers either can't see all the traffic, rely on the same port- and protocol-based classification scheme that has failed the legacy firewall, or provide coverage only for a limited set of applications. Policy management is an even greater problem given that access control rules and inspection requirements are spread among several consoles and involve multiple policy models. And performance is still an issue as well, at least in terms of having a relatively high aggregate latency.

Then comes the kicker: device sprawl. As one "solution" after another is added to the network, the device count, degree of complexity, and total cost of ownership all continue to rise. Capital costs for the products themselves and all the supporting infrastructure that is required are joined by a substantial collection of recurring operational expenditures, including support/maintenance contracts, content subscriptions, and facilities costs (power, cooling, and floor space) — not to mention an array of "soft" costs such as those pertaining to IT productivity, training, and vendor management. The result is an unwieldy, ineffective, and costly endeavor that is simply not sustainable.

Traditional IPSs Are a Poor Match for Today's Threats

IPSs detect and block attacks focused on vulnerabilities that exist in systems and applications. Unlike intrusion detection systems (IDSs), which focus only on alerting, IPS systems are intended to be deployed in-line to actively block attacks as they're detected.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

One of the core capabilities of an IPS is the ability to decode protocols to more accurately apply signatures. This allows IPS signatures to be applied to very specific portions of traffic, thereby reducing the percentage of false positives that were often experienced with signature-only systems. It's important to note that most IPS offerings will use port and protocol as the first pass of traffic classification, which, given the evasive characteristics of today's applications, may lead to an erroneous identification of the application. And because IPSs are focused mainly on attacks, they're typically deployed in conjunction with a firewall as a separate appliance or as a combination firewall and IPS.

IPSs are designed to stop threats using a "find it and kill it" approach. They aren't designed to control applications. But even for stopping threats, IPSs have their flaws.

WARNING

Traditional IPSs don't focus on enforcing outbound user access control and inbound malware prevention during HTTP and Hypertext Transfer Protocol Secure (HTTPS) browsing over the Internet. These threat vectors need to be addressed using an integrated approach (that is, URL filtering in a next-generation firewall).

Given the new application and threat landscape, organizations are also reexamining traditional IPSs. The major IPS vendors are struggling to differentiate across several basic elements of IPSs:

- **Server and data center protection:** There are only a handful of detection and prevention techniques, and most IPS products support them all. These techniques include protocol anomaly detection, stateful pattern matching, statistical anomaly detection, heuristic analysis, blocking of invalid or malformed packets, and IP defragmentation and TCP reassembly (for anti-evasion). Most IPS vendors also use vulnerability-facing signatures (as opposed to exploit-facing signatures) and turn off server-to-client protection to improve performance.
- **Research and support:** This comes down to how much actual research vendors are doing, and how quickly they can respond to help enterprises protect against new attacks and vulnerabilities. Much is made of the efforts of the research teams of IPS vendors, and while there certainly are differences, much of the research is outsourced to a few industry

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

research stalwarts. The other aspect is critical — regardless of who does the research: Can the vendor deliver timely updates to protect customers from new and emerging threats?

- **Performance:** Organizations are clearly sensitized to IPS performance issues. The introduction of traffic/application latency and bandwidth/performance are major concerns that cause enterprises to deploy "out-of-band" IPS. Clearly, being able to keep up with enterprise expectations for throughput and latency is top of mind for many customers.

As defenses mature, however, attackers evolve. Given that intrusion detection and prevention systems, like firewalls, are based on legacy techniques that are relatively well understood, new attacks are able to exploit well-known weak spots, including

- **Application-borne threats:** Threat developers are using applications, both as targets and as transmission vectors. Applications provide fertile ground for both methods. Some application-borne threats are well understood (for example, many of the threats that move across social networks); others are not. Regardless, attackers find it far easier to piggyback on applications and start their attacks with the client. For example, Koobface is a worm that targets users on social media sites such as Facebook and Skype, and is used to steal personal information such as passwords and banking information. CryptoLocker is ransomware that is often propagated via email and encrypts certain files on a victim's computer and requires payment of a ransom (usually in Bitcoin) to decrypt the files.
- **Encrypted threat vectors:** The other important technique that threats employ is encryption. Although security researchers have warned for years that encryption can be used by various threats, encrypted attacks still need a conduit — enter user-centric applications. Users are easily duped into clicking encrypted links (too many users think that HTTPS means "safe"), which can send encrypted threats sailing through enterprise defenses. This is increasingly simple on social networks, where the level of trust is extremely high. The other closely related vector is obfuscation via compression — traditional IPSs can't decompress and, thus, can't scan compressed content.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

A WORD ON DATA LEAKS

Some of the biggest information-security news stories over the past few years involve the leaking of confidential or sensitive organizational data via applications (for example, U.S. federal government agencies and contractors, pharmaceuticals, and retailers). In most cases, the applications that the data leaked across were expressly forbidden — unfortunately, their policies couldn't be enforced with traditional firewalls and IPSs, or alerts (that required manual response) were lost in a sea of information. Given these high-profile security breaches, it's no wonder that organizations are starting to look for a better solution to help protect against such embarrassing incidents.

A common theme here is the level of control needed to prevent these newer threats — controlling applications and content, decrypting SSL/TLS, unzipping content to look for threats — all of which goes well beyond what an IPS traditionally does. A major limitation of an IPS, despite all the work to transition from IDSs, is that IPSs remain a negative security model and are architected as such. Put more simply, an IPS relies on a "find it and kill it" model — which doesn't work very well for the types of control necessary to deal with many of these new threats that move over applications. Nor does it lend itself to an architecture and platform capable of decrypting and classifying all traffic.

A positive security model operates by expressly allowing all communications that are known to be benign, appropriate, or necessary, and excluding everything else. A negative security model operates by seeking to classify only undesirable communications and content, and employing countermeasures for those that are known to be bad.

UTM Only Makes What Is Broken Cheaper

Unified threat management (UTM) devices are another approach to modern security challenges that are nonetheless based on traditional techniques. UTM solutions were born as security vendors began bolting intrusion prevention and antivirus add-ons to

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

their stateful firewalls in an effort to reduce the cost of deployment. UTM products don't perform their functions any better than stand-alone devices. Instead, they provide convenience to the customer by consolidating multiple functions into one device. Unfortunately, UTMs have a reputation for being inaccurate, hard to manage, and performing poorly when services are enabled, relegating them to environments where the value of device consolidation outweighs the downside of lost functionality, manageability, or performance.

The primary advantage of the UTM solution is that it typically does a reasonable job of addressing the issues associated with device sprawl. Instead of having all the "helper" countermeasures deployed as separate devices, with UTM they all come in one physical package.

But so what? The result is really no different than the bolted-on approach and, therefore, exhibits the same deficiencies. Inadequate application classification and resulting blind spots in the inspections that are performed remain fundamental problems, while performance and policy management issues are compounded even further based on having to account for multiple additional countermeasures instead of just one.

It's Time for a Truly Integrated Approach

Traditional port-based firewalls really don't provide value anymore — not in a world where network boundaries are disintegrating and Internet applications are exploding.

But you already know that, which is why you've been forced to make up for their glaring deficiencies with more specialized appliances — IPSs, proxies, antivirus, anti-spyware, URL filtering, and more. Sure, these tools add some incremental value, but it's getting harder to justify their additional cost and complexity — especially during challenging economic times.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

WARNING

More security appliances don't necessarily mean a more secure environment. In fact, the complexity and inconsistency associated with such an approach can actually be a detriment to your organization's security strategy.

Clearly, such a strategy doesn't scale. More important, none of these additional products give you the visibility and control you need over the applications running on your network.

It's time to address the core problem. It's time to make the firewall the visibility and enforcement point for a modern security platform. After all, firewalls deployed at key locations in the network are really the best way to gain visibility and control over what enters and leaves the network.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Identifying applications, users, and content
- Comparing performance between next-generation and legacy firewall architectures
- Recognizing the security and business benefits of next-generation firewalls

Chapter 4
Solving the Problem with Next-Generation Firewalls

Network security in most enterprises is fragmented and broken, exposing them to unwanted business risks and ever-increasing costs. Traditional network security solutions have failed to keep pace with changes to applications, threats, and the network landscape. Furthermore, the remedies put forth to compensate for their deficiencies have proven to be ineffective. It's time to reinvent network security.

This chapter is about the next-generation firewall (NGFW): what it is, what it isn't, and how it can benefit your organization.

The Next-Generation Firewall

To restore the firewall as the cornerstone of enterprise network security, NGFWs "fix the problem at its core." NGFWs classify an organization's network traffic by the application's identity in order to grant access to users and provide visibility and control of

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

all types of applications to admins, including web applications, Software as a Service (SaaS) applications, and legacy applications. The essential functional requirements for an effective NGFW include the ability to:

- Identify applications regardless of port, protocol, evasive techniques, or Secure Sockets Layer (SSL) encryption before doing anything else
- Provide granular visibility and policy-based control over applications, including individual application functions
- Accurately identify users and subsequently use identity information as an attribute for policy control
- Provide real-time protection against a wide array of threats, including those operating at the application layer
- Integrate traditional firewall and network intrusion prevention capabilities
- Support in-line deployments with negligible performance degradation

REMEMBER

Typical capabilities of traditional firewalls include packet filtering, network- and port-address translation, stateful inspection, and virtual private network (VPN) support. Typical intrusion prevention capabilities include vulnerability- and threat-facing signatures, and heuristics.

The key to NGFWs is the ability to do everything a traditional firewall does on top of the advanced capabilities that combine innovative identification technologies, high-performance, and additional foundational features to yield an enterprise-class solution.

Application identification

The first step in application identification is to establish the port and protocol. Next, robust application identification and inspection enables granular control of the flow of sessions through a firewall based on applications that are being used (see Figure 4-1).

Most enterprise web traffic is now encrypted, and attackers exploit encryption to hide threats from security devices. This means even businesses with mature, comprehensive security measures in place can be breached if they aren't monitoring encrypted traffic.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

The ability to decrypt SSL and other encrypted traffic (such as Secure Shell [SSH]) is a foundational security function of NGFWs. Key capabilities include recognition and decryption on any port (inbound or outbound), policy control over decryption, and the necessary hardware and software elements to perform decryption across tens of thousands of simultaneous SSL connections with predictable performance.

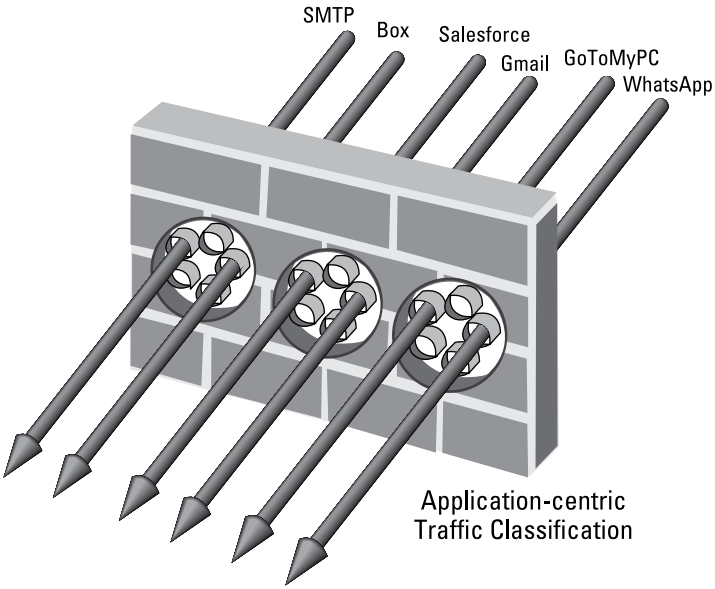


FIGURE 4-1: Application-centric traffic classification identifies specific applications flowing across the network, irrespective of the port and protocol in use.

Positive application identification is the traffic classification engine at the heart of NGFWs. It requires a multifaceted approach to determine the identity of applications on the network, regardless of port, protocol, encryption, or evasive tactics. Application identification techniques used in NGFWs (see Figure 4-2) include

- **Application protocol detection and decryption:** Determines the application protocol (for example, Hypertext Transfer Protocol [HTTP]) and, if SSL/Transport Layer Security (TLS) is in use, it decrypts the traffic so that it can be analyzed further. Traffic is re-encrypted after all the NGFW technologies have an opportunity to operate.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- **Application protocol decoding:** Determines whether the initially detected application protocol is the "real one," or if it's being used as a tunnel to hide the actual application (for example, Tor might be inside of Hypertext Transfer Protocol Secure [HTTPS]).
- **Application signatures:** Context-based signatures look for unique properties and transaction characteristics to correctly identify the application regardless of the port and protocol being used. This includes the ability to detect specific functions within applications (such as file transfers within SaaS applications).
- **Heuristics:** For traffic that eludes identification by signature analysis, heuristic (or behavioral) analyses are applied — enabling identification of any suspicious applications, such as peer-to-peer (P2P) or Voice over Internet Protocol (VoIP) tools that use proprietary encryption.

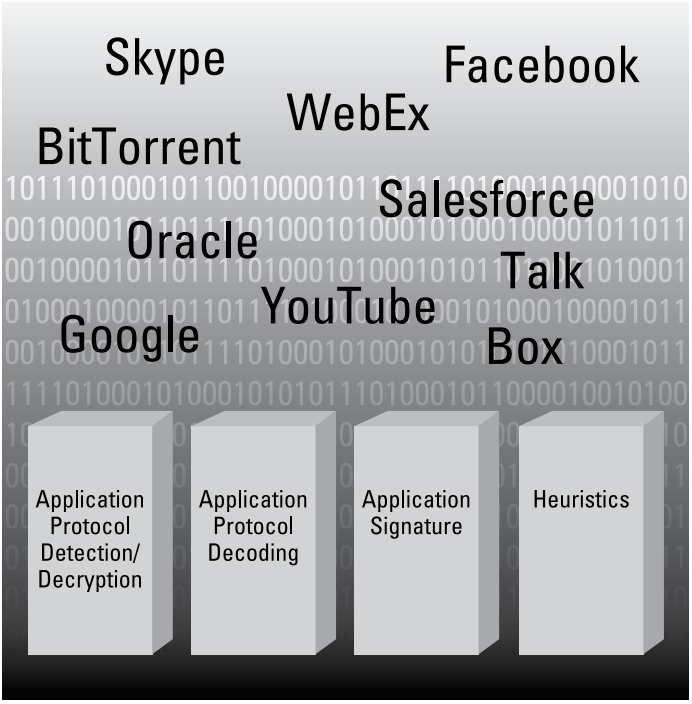


FIGURE 4-2: NGFW techniques used to identify applications regardless of port, protocol, evasive tactic, or SSL encryption.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

TIP

Having the technology to accurately identify applications is important but understanding the security implications of an application so that informed policy decisions can be made is equally important. Look for an NGFW solution that includes information about each application, and its behaviors and risks, to provide IT administrators with application knowledge such as known vulnerabilities, ability to evade detection, file transfer capabilities, bandwidth consumption, malware transmission, and potential for misuse.

User identification

User identification technology links IP addresses to specific user identities, enabling visibility and control of network activity on a per-user basis. Tight integration with Lightweight Directory Access Protocol (LDAP) directories, such as Microsoft Active Directory (AD), supports this objective in two ways:

- It regularly verifies and maintains the user-to-IP address relationship using a combination of login monitoring, end-station polling, and captive portal techniques.
- It communicates with AD to harvest relevant user information, such as role and group assignments.

These details are then available to

- Gain visibility into who specifically is responsible for all application, content, and threat traffic on the network, including users on mobile devices, working remotely, or located in branch offices
- Enable the use of identity as a variable within access control policies
- Facilitate troubleshooting/incident response and reporting

User identification is also an important NGFW capability to help prevent credential theft and abuse. The majority of network breaches today involve stolen credentials that attackers use to simply logon to the network (rather than hacking in) and elevate privileges leveraging other stolen credentials once inside the network.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

With user identification, IT departments get another powerful mechanism to help control the use of applications in an intelligent manner. For example, a remote access application that would otherwise be blocked because of its risky nature can be enabled for individuals or groups that have a legitimate need to use it, such as IT administrators (see Figure 4-3).

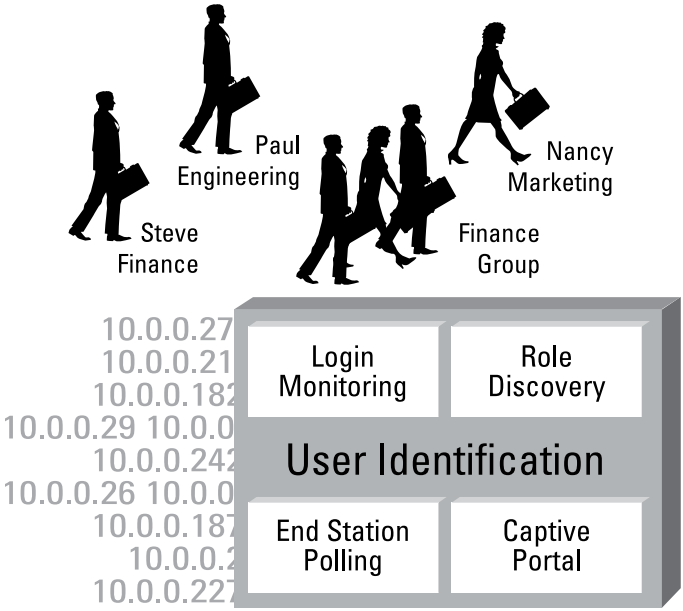


FIGURE 4-3: User identification integrates enterprise directories for user-based policies, reporting, and forensics.

Content identification

Content identification infuses NGFWs with capabilities previously unheard of in enterprise firewalls:

- **Threat prevention:** This component prevents malware and exploits from penetrating the network, regardless of the application traffic in which they are hiding.
  - *Application decoder:* Pre-processes data streams and inspects for specific threat identifiers.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- *Stream-based malware scanning:* Scanning traffic as soon as the first packets of a file are received — as opposed to waiting until the entire file is in memory — maximizes throughput and minimizes latency.
- *Uniform threat signature format:* Performance is enhanced by avoiding the need to use separate scanning engines for each type of threat. Viruses, command-and-control (C2) communications, and vulnerability exploits can all be detected in a single pass.
- *Vulnerability attack protection:* Similar to the functionality provided in IPS devices, protocol anomaly, behavior anomaly, and heuristic detection mechanisms are used for protection from known and unknown threats.
- *Leveraging cloud-based intelligence:* For content that's unknown, the ability to send to a cloud-based security service ("sandboxing") for rapid analysis and a "verdict" that the firewall can then use.

» **URL filtering:** URL filtering is a tool used to classify content. An integrated, on-box URL database allows administrators to monitor and control web surfing activities of employees and guest users. Employed in conjunction with user identification, web usage policies can even be set on a per-user basis, further safeguarding the enterprise from an array of legal, regulatory, and productivity-related risks.

» **File and data filtering:** Taking advantage of in-depth application inspection, file and data filtering enables enforcement of policies that reduce the risk of unauthorized information transfer, or malware propagation. Capabilities include the ability to block files by their actual type (not based on just their extension), and the ability to control the transfer of sensitive data patterns like credit card numbers. Granular policies enable organizations to bypass decryption of certain sensitive data, such as data to and from a known financial institution, if required by security and/or privacy compliance mandates. This complements the granularity of application identification, which offers the ability to control file transfer within an individual application.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

With content identification, IT departments gain the ability to stop threats, reduce inappropriate use of the Internet, and help prevent data leaks — all without having to invest in a pile of additional products that cause appliance sprawl, and that still don't work well because of their lack of integration (see Figure 4-4).

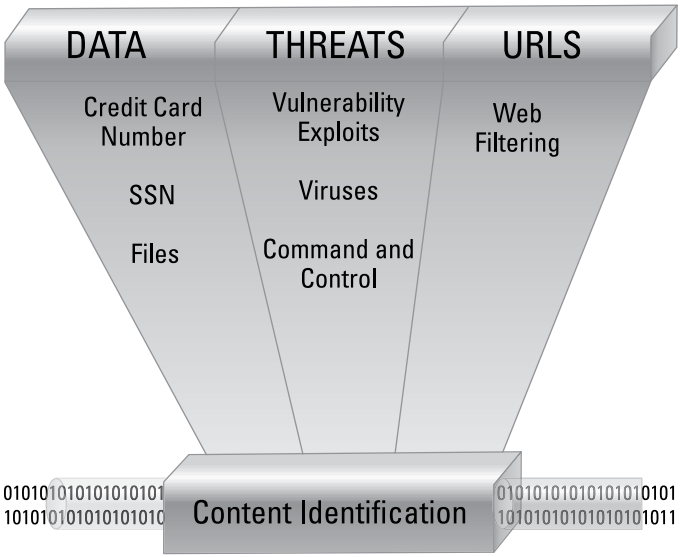


FIGURE 4-4: Content identification unifies content scanning for threats, confidential data, and URL filtering.

Policy control

Identifying the applications in use (application identification), who is using them (user identification), and what they're using them for (content identification) are important first steps in learning about the traffic traversing the network. Learning what the application does, the ports it uses, its underlying technology, and its behavior are the next steps toward making an informed decision about how to treat the application.

When a complete picture of usage is gained, organizations can apply policies with a range of responses that are more fine-grained and appropriate than simply "allow" or "deny" — the only options available in traditional port-based firewalls. This is

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

made possible by the combination of application, user, and content identification, and the positive security model of NGFWs. Traditional port-based firewalls have the security model, but lack intelligence. Other security devices may have some of the intelligence, but not the security model.

REMEMBER

In a positive security model, what is allowed is explicitly defined ("whitelisted") and everything else is rejected. In a negative security model, what is not allowed is explicitly defined ("blacklisted") and everything else is allowed.

Examples of policy control options in NGFWs include

- Allow or deny
- Allow but scan for exploits, viruses, and other threats
- Allow based on device, users, or groups
- Decrypt and inspect
- Apply traffic shaping through Quality of Service (QoS)
- Apply policy-based forwarding
- Allow certain application functions
- Allow (or prevent) certain types of file transfer
- Any combination of the aforementioned

High-performance architecture

Having a comprehensive suite of application awareness and content inspection capabilities is of little value if IT administrators are unable to fully engage them due to performance constraints. So, it's important to select an NGFW that is designed from the start to deliver high performance. The issue is not just that these capabilities are inherently resource intensive. There's also the tremendous traffic volume confronting today's security infrastructure, not to mention the latency sensitivity of many applications. Rated throughput and reasonable latency should be sustainable under heavy loads, even when all application and threat inspection features are engaged simultaneously — which is the ideal configuration from a security perspective.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

For traditional security products, especially those with bolted-on capabilities, each high-level security function is performed independently. This multi-pass approach requires low-level packet processing routines to be repeated numerous times. System resources are used inefficiently and significant latency is introduced (see Figure 4-5).

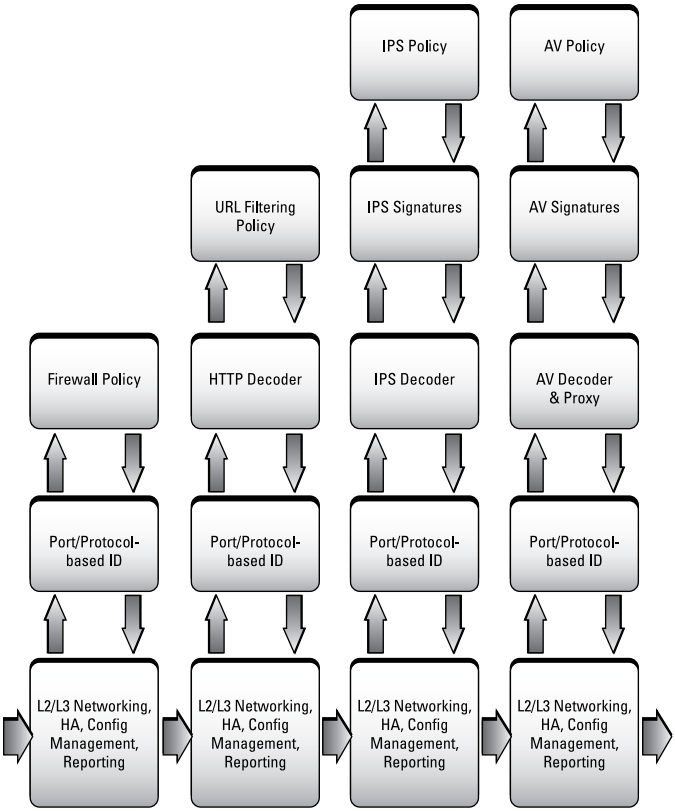


FIGURE 4-5: Legacy multi-pass architectures.

In contrast, an NGFW that uses a single-pass architecture eliminates repetitive handling of packets, reducing the burden placed on hardware and minimizing latency. Separate data and control planes help provide an enterprise-class solution (see Figure 4-6).

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

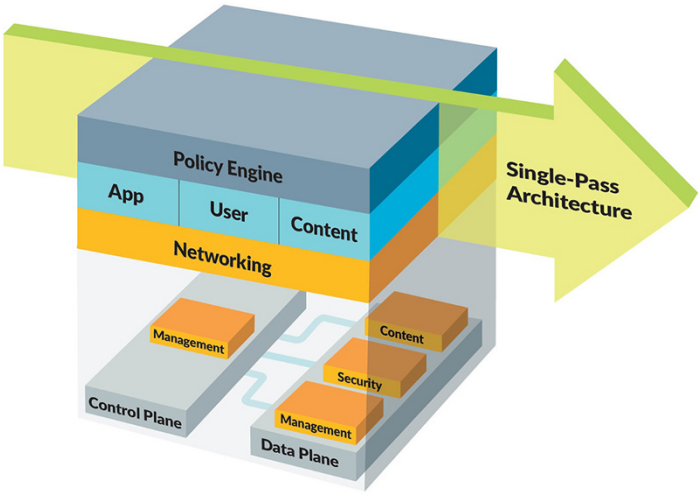


FIGURE 4-6: Single-pass architecture and separate control and data planes provide performance and availability.

What a Next-Generation Firewall Isn't

There are many network-based security products available that perform functions similar to an NGFW, but they aren't the same thing. Here are some examples:

- **Unified threat management (UTM):** UTM appliances host multiple security functions, such as port-based firewall capabilities and basic intrusion prevention. UTM solutions suffer from inferior security capability because not all classification knowledge (for example, the application) is shared with all enforcement options — limiting the flexibility and precision of security policies.
- **Proxy-based products:** Proxies (both firewall and caching) sit between source and destination, intercepting traffic and inspecting it by terminating the application session and reinitiating it to the target destination. The proxy establishes the connection with the destination, on behalf of the client, hiding computers on the network behind the proxy. However, only a limited number of applications can be supported because each individual application has to have its own proxy.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

REMEMBER

- **Web application firewalls (WAFs):** A WAF is designed to look at web applications, monitoring them for security issues that may arise due to possible coding errors. WAFs look only at Layer 7, rather than inspecting the entire Open Systems Interconnection (OSI) stack.

  WAFs protect applications, and NGFWs protect networks.
- **Vulnerability and patch management:** Vulnerability and patch management solutions scan hosts for known vulnerabilities in software and operating systems, verify that patches and updates are installed, and correct the identified vulnerability. This is not a function of NGFWs.
- **Data loss prevention (DLP):** These solutions prevent transmission of data that matches an identified pattern (such as credit card numbers). These solutions are implemented for network functions with no real-time requirements regarding speed and latency.
- **Secure web gateways:** These solutions use URL categorization to enforce policies regarding user access to websites and block malware propagated by malicious websites. Compared to NGFWs, these solutions have limited capabilities and are easily circumvented by users.
- **Secure messaging gateways:** These include spam filters and IM gateways, and provide antispam and anti-phishing protection, antivirus scanning, attachment filtering, content filtering, data loss prevention, and policy compliance and reporting. Unlike NGFWs, these functions are not performed in real-time and are used for applications like email, which is less latency sensitive.

Benefits of Next-Generation Firewalls

NGFWs produce numerous benefits over traditional network security infrastructures and solutions:

- **Visibility and control:** The enhanced visibility and control provided by NGFWs enable enterprises to focus on business-relevant elements such as applications, users, and content for policy controls, instead of having to rely on nebulous and

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

misleading attributes like ports and protocols, and to better and more thoroughly manage risks and achieve compliance, while providing threat prevention for allowed applications.

- **Safe enablement:** Achieve comprehensive coverage — by providing a consistent set of protection and enablement capabilities for all users, regardless of their location.
- **Simplification:** Reduce complexity of the network security and its administration — by obviating the need for numerous stand-alone products. This consolidation reduces hard capital costs, as well as ongoing "hard" operational expenses, such as support, maintenance, and software subscriptions, power and heating, ventilation, and air condition (HVAC), and "soft" operational expenses, such as training and management.
- **IT and business alignment:** Enable IT to confidently say "yes" to the applications needed to best support the business — by giving them the ability to identify and granularly control applications while protecting against a broad array of threats. This includes the ability for IT to add security rules "in stages" — actively investigating traffic that is unknown (based on advanced visibility) and then adding security rules as appropriate.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Implementing employee, desktop, and network controls
- Asking the right questions to help you choose the best solution
- Designing your network for optimum performance and security

Chapter 5
Deploying Next-Generation Firewalls

Far too often, technical solutions are implemented without considering the implications for an organization's overall security strategy. To avoid this mistake, it's important to ensure that your policies are up to date and the technology solutions you're considering support a comprehensive security strategy. It's also important to have a clear understanding of your organization's requirements.

This chapter describes the different types of controls that must be considered in an organization's security policies and provides specific examples of technical requirements you need to explore as you define your requirements and develop a request for proposal (RFP) for your vendors. Finally, it covers the importance of properly segmenting your network and sensitive data and how to address mobile users.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Defining Your Requirements and Developing a Request for Proposal

After creating or updating your organization's security policies, it's time to define your organization's requirements for a next-generation firewall (NGFW) solution. At a very high level, this includes doing your due diligence on the vendors you're considering. You should be asking questions about your potential vendors, such as the following:

- What is the company's vision and how well does it execute on that vision? Does it address the current threat landscape, as well as support the IT architectures desired (for example, virtualization and cloud)?
- How innovative is the company? How well regarded is the company — in terms of both technology and its dedication to customer success?
- What is the company's culture?
- What is its development process? What is its quality assurance process?
- What is the size and financial condition of the company?
- Is the company a potential acquisition target? If so, is it more likely to be acquired in order to quickly gain an edge because of its innovation and proprietary technology, or to kill off a competitor?
- How large is its installed customer base?
- Does it have other customers (perhaps even competitors) that are in a similar industry as your own organization?
- Does it have any reference accounts or customer success stories to share?

Next, define your organization's technical requirements. Fortunately, you don't necessarily have to reinvent the wheel here. Begin by taking a look at your organization's security policies (see the previous section) to see what capabilities will be needed in order to implement and support those policies.

There are also plenty of examples of firewall and network security requirements practically everywhere. In fact, most regulatory compliance requirements relating to data protection are based on information security best practices. Even if your organization

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

isn't subject to any of these regulations, using them for guidance isn't necessarily a bad thing. For example, the Payment Card Industry (PCI) Data Security Standard (DSS), which is applicable to every organization that processes a credit or debit card, defines several firewall requirements, all of which can easily be modified and incorporated into a formal RFP for your organization.

Drilling down into specific feature requirements, your RFP should address several requirements, covered in the following sections.

Application identification

Describe how the gateway will accurately identify applications and the mechanisms used to classify applications:

- Is identification based on intrusion prevention system (IPS) or deep packet inspection (DPI) technology? If so, how are accuracy, completeness, and performance issues addressed when scanning network traffic?
- How is the traffic classification mechanism differentiated from other vendors?
- How are unknown applications handled?
- Are custom application signatures supported?
- How is Secure Sockets Layer (SSL)/Transport Layer Security (TLS)–encrypted traffic identified, inspected, and controlled?
- How do the SSL controls delineate between personal traffic (such as banking, shopping, and health) and nonpersonal traffic?
- Are applications identified in a business-relevant way, and what is the process for updating the application database (for example, software upgrade or dynamic update)?
- If a new application is needed, what is the process for adding it to the device?
- Can an end user submit an application for identification and analysis and/or define custom applications?
- Does the product support URL filtering? Describe the URL filtering database. Is the database located on the device or on another device?
- Describe/list any other security functions that can leverage the application information collected, including drill-down details and user visibility features.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Application policy control

Describe the process for implementing policy-based application controls, all application policy control parameters (such as user, Internet Protocol [IP] address, date/time), and how they can be used:

- Can policy controls be implemented for all applications identified?
- Can policy controls be implemented for specific users or groups?
- Can all policy controls be based on the application identified (for example, file blocking)?
- How are remote access environments (for example, Citrix and Terminal Services) supported?
- Can the solution perform traditional firewall-based access controls?
- Can policy controls be implemented from a single management interface?
- Are users warned when they try to access a URL or application that violates policy?

Threat prevention

Describe the intrusion prevention features and antivirus engine:

- List the types of threats that can be blocked. List the file types that can be blocked.
- Is data filtering supported?
- Can the threat prevention engine scan inside SSL/TLS-encrypted traffic? Compressed traffic?
- Can the use of strong cipher and encryption protocol versions be enforced?
- Can combinations of events (indicators of compromise [IOCs]) be recognized so that security operators can be notified of possible issues (for example, compromised endpoints)?

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Management

Describe the management capabilities and visibility tools that enable a clear picture of the traffic on the network:

- Does device management require a separate server or device?
- Are application policy controls, firewall policy controls, and threat prevention features all enabled from the same policy editor?
- What tools provide a summary view of the applications, threats, and URLs on the network?
- Describe any log visualization tools.
- Are reporting tools available to understand how the network is being used and to highlight changes in network usage?
- Describe the logging and reporting capabilities of the solution.
- Describe how management access is ensured when the device is under heavy traffic load.
- Are any central management tools available?

Networking

Describe the network integration and implementation capabilities:

- Describe any Layer 2 or Layer 3 capabilities.
- Are 802.1q virtual local area networks (VLANs) supported? What is the VLAN capacity?
- Is dynamic routing supported (for example, Open Shortest Path First [OSPF], Border Gateway Protocol [BGP], and Routing Information Protocol [RIP])?
- Is equal-cost multipath (ECMP) routing supported for performance and reliability?
- Describe any Quality of Service (QoS) or traffic shaping features.
- Is Internet Protocol version 6 (IPv6) supported?
- Are Internet Protocol Security (IPsec) virtual private networks (VPNs) supported? SSL VPNs?
- What deployment options are available (for example, in-line, tap, passive)?
- Describe any high availability (HA) capabilities.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Hardware

Is the solution software-based, an original equipment manufacturer (OEM) server, or a purpose-built appliance? Describe the architecture along with any performance implications.

IT solution

Does the NGFW need to integrate with other systems? This can be other security subsystems (for example, security information and event management [SIEM]) or a larger IT architecture (for example, cloud and software-defined networking [SDN]). What application programming interfaces (APIs) and features are present to support the integration?

Safe Enablement through Smart Policies

Enablement is first and foremost about education and knowledge of applications, behavior, risks, and users. In the case of Software as a Service (SaaS) and web applications, the users have long since decided on the benefits, although there continue to be opportunities for education on the choice of the best application for the job. The role of IT is that of an advisor and mentor, advising users about risks and behaviors and guiding them regarding which of the array of available applications may be best at solving their requirements. But enablement is also about raising the awareness of the risks associated with applications. For IT to be relevant, it needs to evaluate and adopt SaaS and web applications wholeheartedly and without prejudice. When that's achieved, IT can successfully educate the users on all the risks associated with the use of those applications.

It comes down to using the right tool for the job and being smart about it. For example, in a heavily regulated environment such as stock trading, the use of instant messaging may be prone to retention and auditability rules. The role of IT is to educate the traders on the implications of each of the tools, participate in the development of the use policy, and subsequently monitor and enforce its use. In this example, that policy could prevent the traders from using Facebook Messenger for instant messaging but enable Microsoft Teams for that use instead.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

REMEMBER

Governance and its management counterpart work best if they're based on a set of smart corporate policies that are developed by the four major stakeholders in the application landscape: IT, HR, executive management, and the users. Clearly, IT has a role to play, but it can't be the strictly defined role that IT so often plays, nor can IT be lax about its role as the enabler and governor of applications and technology.

TIP

If application controls are going to be implemented and enforced, they should be part of the overarching corporate security policy. As part of the process of implementing an application control policy, IT should make a concerted effort to learn about all applications that are being used in the organization, including SaaS and web applications. This includes embracing them for all their intended purposes and, if needed, proactively installing them or enabling them in a lab environment to see how they act. Peer discussions, message boards, blogs, and developer communities are valuable sources of information.

Employee controls

Most companies have some type of application usage policy, outlining which applications are allowed and which are prohibited. Every employee is expected to understand the contents of this policy and the ramifications of not complying with it, but there are a number of unanswered questions, including the following:

- Given the increasing number of "bad" applications, how will an employee know which applications are allowed and which are prohibited?
- How is the list of unapproved applications updated, and who ensures employees know the list has changed?
- What constitutes a policy violation?
- What are the ramifications of policy violations — firing or a reprimand?

The development of policy guidelines is often challenging because tension between risk and reward has polarized opinions about what should be allowed and what should be prohibited. At the core of the issue is the fact that the two organizational groups that are typically involved in policy development — IT security and HR — have largely been sidelined during adoption of new technologies. To build a policy for safe use after new technologies and applications have been implemented is no easy task.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Documented employee policies need to be a key piece of the application control puzzle, but employee controls as a stand-alone mechanism will remain largely ineffective for safe enablement of applications.

Desktop controls

Desktop controls present IT departments with significant challenges. Careful consideration should be applied to the granularity of the desktop controls and the impact on employee productivity. As with employee policies, desktop controls are a key piece to the safe enablement of applications in the enterprise and, if used alone, will be ineffective for several reasons.

The drastic step of desktop lockdown to keep users from installing their own applications is a task that is easier said than done:

- Laptops connecting remotely, Internet downloads, USB drives, and email are all means of installing applications that may or may not be approved.
- Removing administrative rights completely has proven to be difficult to implement and, in some cases, limits end-user capabilities.
- USB drives are now capable of running applications, so an application can, in effect, be accessed after the network admission is granted.

Desktop controls can complement the documented employee policies as a means to safely enable applications.

Network controls

At the network level, what is needed is a means to identify all applications and block or control them. By implementing network-level controls, IT is able to minimize the possibility of threats and disruptions stemming from the use of applications.

Several possible control mechanisms can be used at the network level, each of which has certain drawbacks that reduce its effectiveness:

- **Stateful firewalls:** Stateful firewalls can be used as a first line of defense, providing coarse filtering of traffic and segmenting the network into different password-protected zones.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

One drawback to stateful firewalls is that they use protocol and ports to identify and control what gets in and out of the network. This port-centric design is relatively ineffective when faced with applications that hop from port to port until they find an open connection to the network.

- **IPSs:** An IPS added to a firewall deployment enhances the network threat prevention capability by looking at a subset of traffic and blocking known threats or bad applications. IPS offerings lack the breadth of applications and the performance required to look at all traffic across all ports and, as such, cannot be considered a full solution.
- **Proxy solutions:** Proxy solutions are another means of traffic control, but they look at a limited set of applications or protocols and, as such, only see a partial set of the traffic that needs to be monitored. So an application will merely see a port blocked by a proxy and hop over to the next one that is open. By design, proxies need to mimic the application they're trying to control, so they struggle with updates to existing applications, as well as development of proxies for new applications. A final issue that plagues proxy solutions is throughput performance brought on by how the proxy terminates the application, and then forwards it on to its destination.

WARNING

Managing a firewall and IPS combination is usually a cumbersome task, requiring different management interfaces pointed at separate policy tables. Simply put, the current bolt-on solutions don't have the accuracy, policy, or performance to solve today's application visibility and control requirements.

The challenge with all these network controls is that they don't have the ability to identify all applications; they look at only a portion of the traffic and suffer from performance issues.

Data Centers Are Highly Dynamic and Workloads Are Everywhere

Data is everywhere and modern data centers are now highly distributed. New technologies such as software-defined networking (SDN), virtualization, and hyperconvergence are increasingly being adopted in the data center. Applications are also changing

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

and influencing the design and evolution of modern data centers. Application workloads move dynamically and constantly across multiple data centers and public and private cloud environments. Finally, new application technologies like containers and microservices are changing how modern, cloud-native applications are designed and built, while continuous integration (CI) and continuous delivery (CD) pipelines enable DevOps teams to deploy new applications more quickly.

Another important shift in the modern data center is the flow of network traffic. In the past, most network traffic was client–server based and moved in and out of the data center (north–south) across a firewall boundary (the *perimeter*). Today, most network traffic is server to server (east–west) and often times doesn't traverse a traditional firewall boundary. Organizations sometimes "hairpin" (or backhaul) east-west traffic from the data center network fabric to a perimeter firewall for inspection and policy enforcement, but this technique introduces latency and wastes bandwidth.

It's important to design your network to maximize performance and efficiency. Properly deploying an NGFW in the most optimal location(s) on your network is no less important. To maximize deployment flexibility, you need an NGFW that is available in a variety of hardware- and software-based form factors to meet different performance requirements in on-premises data centers, branch locations, and cloud environments.

The explosive growth of virtualization and cloud computing introduces new security challenges that are difficult or impossible for legacy firewalls to manage effectively due to inconsistent functionality, disparate management, and a lack of integration points within the virtualized environment. In order to protect traffic flowing in and out of the data center within your virtualized environments and in the public cloud, an NGFW must support the same functionality in both a hardware and virtualized software form factor.

The dynamic setup and teardown of applications within a virtualized data center exacerbates the challenges of identifying and controlling applications using a port- and IP address–centric approach. An NGFW must provide in-depth integration with the virtualization environment to streamline the creation of application-centric policies as new virtual machines (VMs) and

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

applications are established and taken down. This is the only way to ensure you can support evolving data center architectures with operational flexibility while addressing risk and compliance requirements.

What's needed for modern data center security is an NGFW that provides

- **Deep visibility:** Sees everything across users, devices, networks, and applications
- **Multi-layered segmentation:** Reduces the attack surface across north–south and east–west traffic
- **Automated threat protection.** Uses layered defenses to quickly find threats and stop breaches

REMEMBER

When evaluating an NGFW platform, be sure to also confirm and validate the interaction with the virtualization platform management system, as well as automation and orchestration systems.

Addressing Mobile and Remote Users and Branch Locations

Another technical limitation for traditional firewalls is providing visibility and control for users who are mobile or remote, beyond the perimeter established by enterprise firewalls. The challenge for NGFWs in this case is to deliver a solution that provides the same degree of protection and application enablement received by users on the local network without having to manage a completely independent set of policies.

Another major challenge is to avoid the limitations and disadvantages associated with the current crop of solutions in this area, including the following:

- **Cloud or customer premises equipment (CPE)–based proxies:** Associated web services and products typically focus on a narrow traffic stream (for example, port 80/Hypertext Transfer Protocol [HTTP] only), can have a limited set of services/countermeasures (such as URL or malware filtering only), and — because they rely on a

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

proxy architecture — often have to allow many applications to bypass their filters in order to avoid breaking them.

- **Backhaul via VPN technology:** Whether it's IPsec or SSL-based makes little difference. There is an inevitable bump in latency as client traffic is directed back to one of a few central sites where the VPN gateways are typically located. Of even greater concern, however, is the lack of application visibility and control of the head-end devices that are subsequently used to identify and filter this traffic.

In comparison, a solution that relies on a persistent client that can be installed on demand provides a better alternative. Like the VPN-based approach, remote traffic is sent over a secure tunnel. The difference in this case is that the connection is automatically made to the nearest NGFW — whether it's deployed at one of an organization's hub facilities, out in a regional or branch office location, or as part of a public/private cloud implementation. The latency impact is, thus, minimized, and the user's session is protected and controlled by the full portfolio of application-, user-, and content-oriented identification and inspection technologies — exactly as if the user were operating on the local network instead of remotely. The net result is an easy-to-implement solution that provides remote and mobile users with the same degree of application enablement and protection as their in-office counterparts.

Adopting Best Practices to Maximize Security Effectiveness

Zero Trust is a strategy designed around the concept that users, applications, and data should never be trusted — that their actions should always be verified in an environment. The primary goal of the Zero Trust model is to eliminate inherent trust in a system and prevent attackers from exploiting vulnerabilities hidden in trusted applications. The approach involves limiting the scope of an attack and blocking lateral movement by taking advantage of micro-segmentation based on users, data, and location.

Segmentation is a key concept in the proper design of networks and deployment of firewalls. Although there are many different

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

ways to segment a network, NGFWs bring a unique combination of hardware- and/or software-related segmentation capabilities that enable organizations to isolate key sections of their networks, including data centers, the Internet edge (perimeter), branch locations, and public, private, and hybrid cloud environments. An NGFW can act as a segmentation gateway to enable a Zero Trust architecture.

The concept of security zones, which for purposes of isolating sensitive data or critical network infrastructure (again, for example, a data center), is roughly equivalent to that of network segments (see Figure 5-1). A security zone is a logical container for physical interfaces, virtual local area networks (VLANs), a range of IP addresses, or a combination thereof. Interfaces that are added to each security zone can be configured in Layer 2, Layer 3, or a mixed mode, thereby enabling deployment in a wide range of network environments without requiring network topology modifications.

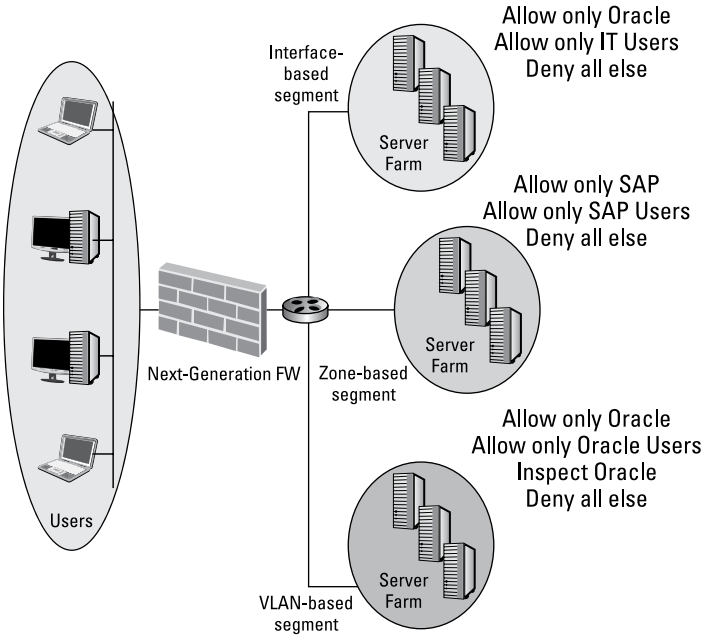


FIGURE 5-1: Network segmentation and security zones.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Many different technologies can be used to segment the network, but when looking at segmentation as a way to isolate the sensitive data or critical infrastructure, several key requirements need to be taken into account:

- **Flexibility:** Segmenting the network for security purposes may sometimes require the modification of the network architecture, a task that most companies will avoid if at all possible. The ability to segment a network using IP address ranges, VLANs, physical interfaces, or a combination thereof, is paramount.
- **Policy-based security:** Policies must be based on the identity of users and the applications in use — not just IP addresses, ports, and protocols. Without knowing and controlling exactly who (users) and what (applications and content) has access within a segment, sensitive data may be exposed to applications and users that can easily bypass controls based on IP addresses, ports, and protocols.
- **Performance:** Segmentation means applying in-depth security policies in a network location that is typically business-critical, high-volume traffic. This means it's critical that the solution delivering the secure segment operate at high speeds with very high session rates and minimal latency.

REMEMBER

An NGFW platform can be used to implement a Zero Trust architecture by enabling secure access for all users irrespective of location, inspecting all traffic, enforcing policies for least-privilege access control, and detecting as well as preventing advanced threats. This significantly reduces the pathways for adversaries to access your most critical data and applications, whether the adversaries are outside or inside your organization.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Identifying critical security functions
- Defining operational requirements
- Ensuring performance, scalability, and extensibility

Chapter 6
Ten Things Your Firewall Must Do

If your organization is still relying on port-based firewalls and bolt-on security point products to secure its network, then you're exposed. It's time to replace these legacy products with a next-generation firewall (NGFW) as the cornerstone of an effective enterprise network security strategy. Here's what your NGFW must be able to do.

Identify Users and Enable Appropriate Access

Employees, customers, and partners connect to different repositories of information within your network, as well as to the Internet. These people and their many devices represent your network's users. It's important for your organization's risk posture that you're able to identify your users beyond their Internet Protocol (IP) addresses, as well as grasp the inherent risks they bring based on the devices they're using — especially when security policies have been circumvented or new threats have been introduced to your network.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

In addition, users are constantly moving to different physical locations and using multiple devices, operating systems, and application versions to access the data they need (see Figure 6-1). IP address subnets are mapped only to physical locations, not to individual users, meaning that when users move around — even within the office — security policies don't follow them.

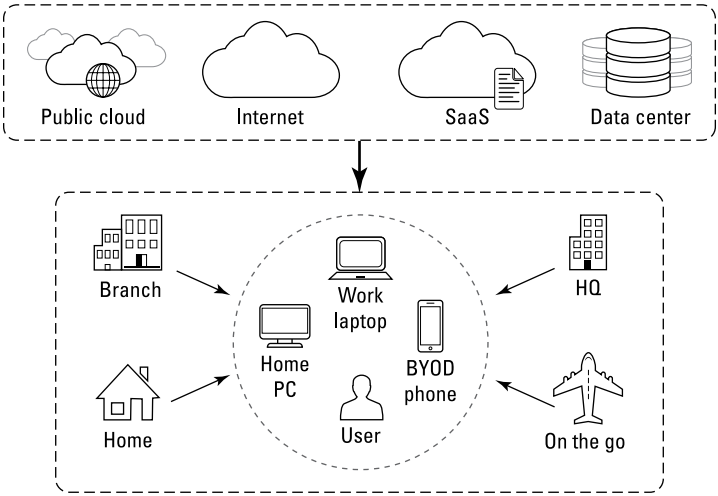


FIGURE 6-1: Users access data from different devices and locations.

User and group information must be directly integrated into the technology platforms that secure modern organizations. Your firewall must be able to pull user identity from multiple sources, including virtual private networks (VPNs), wireless local area network (WLAN) access controllers, directory servers, email servers, and captive portals. Knowing who is using the applications on your network, and who may be spreading a malware threat or transferring files, strengthens security policies and improves incident response times.

The firewall must allow policies to safely enable applications based on users or groups of users, outbound or inbound — for example, by allowing only your IT department to use tools such as Secure Shell (SSH), telnet, and File Transfer Protocol (FTP).

REMEMBER

User-based policies follow users no matter where they go — at headquarters, branch offices, or home — and on whatever devices they use, but user identity goes beyond classifying users for policy reporting.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Prevent Credential Theft and Abuse

Users and their credentials are among the weakest links in an organization's security infrastructure. According to Forrester Research, at least 80 percent of data breaches today involve compromised privileged credentials. With stolen credentials as part of their tool kit, attackers' chances of successfully breaching network defenses go up and their risk of being discovered goes down. Attackers can easily steal credentials through phishing, malware, social engineering, or brute force, and they can even buy credentials on the dark web. Attackers use these credentials to gain access to a network, move laterally, and escalate their privileges for unauthorized access to applications and data.

Traditional enterprise security efforts to prevent credential theft and abuse include

- **User awareness training:** Training about topics such as password security and email phishing is an important first step in mitigating the risk of credential theft through social engineering.
- **Password policies:** Commonly implemented safeguards include periodic password changes, unique passwords that meet minimum length and complexity requirements, and account lockouts.
- **Security point products:** These products attempt to identify known phishing sites and filter email to prevent malware-infected spam and phishing emails from reaching users' inboxes.

The effectiveness of these methods varies widely across different organizations and is somewhat limited in today's threat landscape. Social engineering techniques, particularly phishing campaigns, are becoming increasingly sophisticated and difficult to detect, and it only takes one human mistake to breach a network with stolen credentials. Static passwords are inherently vulnerable, and many password policies lead to unsafe and undesired user behaviors such as writing down passwords and using minor variations of the same passwords across different accounts. Products that check for known bad sites often miss newly created

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

ones and attackers can bypass email filtering altogether by sending malicious links through social media.

Organizations need a firewall with machine-learning-based analytics to identify websites that steal credentials in real time. The firewall should be able to automatically update its rule base and dynamically block newly identified malicious websites and URLs. Still, there will always be new, never-before-seen phishing sites that are treated as "unknown." Your firewall must allow you to prevent corporate credentials from being submitted to unknown sites. The firewall must also allow you to protect sensitive data and applications by enforcing multifactor authentication (MFA). When properly implemented, MFA can all but eliminate the risk of attackers using stolen credentials to access network resources. By integrating with common MFA vendors, your firewall can protect your applications containing sensitive data, including legacy applications.

MFA requires a combination of at least two factors, typically "something you know" (that is, an account name and password) and "something you have" (such as a smartphone associated with the account or a hardware/software token). In a typical MFA implementation, a user submits a valid account name and password and then receives a one-time numeric code via a phone call, text message, or token authenticator. The code is valid only one time and only for a limited time (typically, less than ten minutes). If the user enters the code incorrectly or the code expires, a new code must be sent to the smartphone or token for the user to successfully authenticate because the code is only valid for a single logon. Thus, an attacker would need to physically possess the user's smartphone or token (without the user's knowledge — increasingly unlikely given that so many users' smartphones are physically attached to their hands!) to breach a network using stolen account credentials. Even if the attacker intercepts the one-time code as it's being sent to the user's device, the code is only valid for a single logon session. If the code is being intercepted and used by the attacker in real time during an interactive logon session, the legitimate user will be unable to log on with that code and will, in most cases, resubmit their account credentials to have a new code sent to their device. If MFA is properly implemented, after the new code is used to authenticate, the previous logon session (that was intercepted) would no longer be valid and the attacker's session would be automatically disconnected.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Safely Enable All Applications and Control Functions

Users are increasingly leveraging a variety of applications — including Software as a Service (SaaS) applications from multiple devices and locations — for work-related (as well as personal) purposes. Many applications, such as instant messaging (IM), peer-to-peer (P2P) file sharing, and Voice over Internet Protocol (VoIP), are capable of operating on nonstandard ports or hopping ports. Some of these applications are sanctioned by the organization, others are tolerated, and others are unsanctioned. Users are increasingly savvy enough to force applications to run over nonstandard ports through protocols such as Remote Desktop Protocol (RDP) and SSH, regardless of the organization's policy regarding various applications (sanctioned, tolerated, unsanctioned).

Many applications include rich sets of features and functionalities that provide important business capabilities but may represent different risk profiles. For example, WebEx is a valuable business tool, but using WebEx desktop sharing to take over an employee's desktop from an external source may be an internal or regulatory compliance violation. Gmail and Google Drive are also good examples. When users sign on to their Gmail accounts (which may be allowed), they can easily switch to YouTube or Google Photos (which may not be allowed). Security administrators need the ability to control application usage within the corporate environment by creating policies to allow certain applications and application functions, while denying others.

Your firewall must be able to classify traffic by application on all ports, all the time, by default — and it should create an administrative burden by requiring you to research which applications use which ports in order to configure firewall rules. The firewall must provide complete visibility into application usage along with capabilities to understand and control their use (see Figure 6-2).

For example, it should understand usage of application functions, such as audio streaming, remote access, and posting documents, and be able to enforce granular controls over that usage, such as upload versus download permissions, chat versus file transfer, and so on. This must be done continuously. The concept of "one-and-done" traffic classification is not an option because it

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

ignores the reality that these commonly used applications share sessions and support multiple functions. If a different function or feature is introduced in the session, the firewall must perform a policy check again. Continuous state tracking to understand the different functions each application may support — and its associated risks — is a must for your firewall.

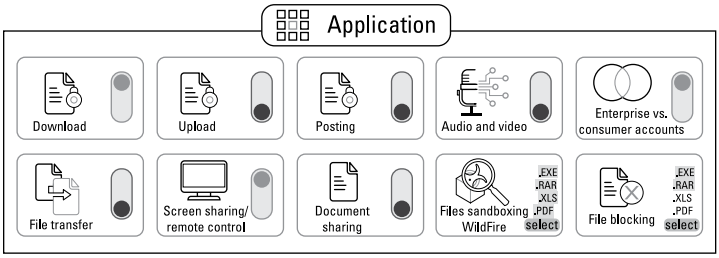


FIGURE 6-2: Control application usage in policy.

Secure Encrypted Traffic

Most web traffic is now Secure Sockets Layer (SSL) encrypted, and attackers exploit encryption to hide threats from security devices. This means even businesses with mature, comprehensive security measures in place can be breached if they aren't monitoring SSL-encrypted traffic. Additionally, SSH is used nearly universally by tech-savvy users to hide non-work-related activity.

The ability to decrypt SSL and SSH is a foundational security function. Key elements to look for include recognition and decryption on any port (inbound or outbound), policy control over decryption, and the necessary hardware and software elements to perform decryption across tens of thousands of simultaneous SSL connections with predictable performance.

Your firewall must also be flexible enough to easily decrypt certain types of encrypted traffic (such as Hypertext Transfer Protocol Secure [HTTPS] from unclassified websites), while other types of encrypted traffic (such as web traffic from known financial services organizations) are not decrypted to comply with security and privacy mandates. A firewall should apply security and load balancing to decrypted flows across multiple stacks of security devices for additional enforcement. This eliminates the need for dedicated SSL offloading and reduces network complexity.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Stop Advanced Threats to Prevent Successful Cyberattacks

Most modern malware (including ransomware variants) uses advanced techniques — such as wrapping malicious payloads in legitimate files or packing files to avoid detection — to transport attacks or exploits through network security devices and tools. As organizations have increasingly deployed virtual sandboxes for dynamic analysis, attackers have evolved their techniques to evade sandbox analysis, for example, by scanning for valid user activity, system configurations, and indicators of specific virtualization technologies. With the growth of the cybercrime, any attacker — novice or advanced — can purchase plug-and-play tools designed to identify and avoid malware analysis environments.

Using integrated security services, your firewall should automatically prevent known threats and automatically analyze and prevent unknown threats. You need a firewall that looks for threats at all points within the attack life cycle (see Figure 6-3), not just when threats first enter your network. Blocking known risky file types or access to malicious URLs before they compromise your network reduces your threat exposure. Your firewall should protect you from known vulnerability exploits, malware, and command-and-control (C2) activity without requiring you to manage or maintain multiple single-function appliances. Signatures should be updated automatically as soon as new malware or threats are encountered to keep your organization and users protected.

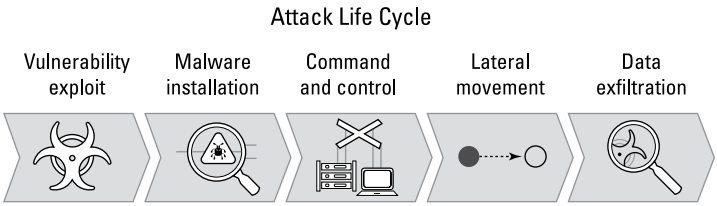


FIGURE 6-3: Your next firewall must detect and prevent threats across the entire attack life cycle.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

A firewall that utilizes multiple methods of analysis to detect unknown threats, including static analysis with machine learning, dynamic analysis, and bare-metal analysis, is capable of high-fidelity, evasion-resistant discovery. Instead of using signatures based on specific attributes, firewalls should use content-based signatures to detect variants, polymorphic and metamorphic malware, or C2 activity. In addition, C2 signatures based on analysis of outbound communication patterns are much more effective protective measures that can scale at machine speed when created automatically. Finally, cloud-delivered security infrastructure is critical for security enforcement. It supports threat detection and prevention at massive scale across your network, endpoints, and clouds in addition to allowing you to tap into an open ecosystem of trusted security innovators.

Protect Your Growing Mobile Workforce

Mobile devices are increasingly used by the growing mobile workforce to connect to business applications, often through public networks and on devices that are vulnerable to advanced threats. Threats and risks increase exponentially when users are off premises because there is no network firewall to prevent attacks against these users and their devices. Cloud and bring-your-own-device (BYOD) trends add to the complexity in addressing these threats and risks. In addition, remote locations and small branch offices often lack consistent security because it's operationally inefficient and costly to ship firewalls to them or backhaul traffic to headquarters.

The mobile workforce and remote locations need access to applications from places far beyond your network. They also need protection from targeted cyberattacks, malicious applications and websites, phishing, C2 traffic, and other unknown threats. This requires consistent security everywhere. Your firewall must enable the required levels of visibility, threat prevention, and security policy enforcement to protect your distributed users and locations by delivering security capabilities from the cloud, without the need to deploy physical hardware.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Extend Security to Your Evolving Cloud Environments

Data and applications are everywhere. According to the RightScale 2018 *State of the Cloud Report,* 81 percent of enterprises use multiple public, private, and/or hybrid clouds — five different clouds on average. Organizations must secure sensitive data in the network and across a variety of cloud platforms, including SaaS environments. Many legacy security tools and techniques, designed for static networks, are ineffective and/or incompatible with cloud-native tools and capabilities. Moreover, native security services from the cloud providers themselves, such as Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure, typically provide only Layer 4 protections and are specific to the cloud provider.

Your organization needs cloud security that extends policy consistently from the network to the cloud, stops malware from accessing and moving laterally (east–west) within the data center and cloud, simplifies management, and minimizes the security policy lag as virtual workloads change. Your firewall must protect the resident applications and data with the same security posture that you may have established on your physical network. To secure multicloud deployments, the firewall must support a variety of cloud and virtualization environments, including all major public cloud providers and virtualized private clouds. The firewall must integrate with native cloud services, such as Amazon Lambda and Azure, and automation tools, such as Ansible and Terraform, to integrate security into your cloud-first development projects.

Centralize Management and Integrate Security Capabilities

Individual point security products typically come with their own management applications. To configure security for each product, security operators must work with different management devices. According to the 2017 *U.S. IT Services Report* from ResearchCorp, nearly 72 percent of organizations use products from three or more separate vendors to secure their network infrastructure. These products are disconnected and cannot share insights.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Organizations also find it challenging to scale firewall onboarding, maintain consistent security policies, and deploy emergency changes across thousands of firewalls. This makes security complex and stretches IT teams to their limits.

REMEMBER

According to a leading analyst firm, 99 percent of firewall breaches will be caused by misconfiguration.

You must be able to operationalize the deployment of consistent, centralized security policies across tens of thousands of firewalls spanning on-premises and cloud deployments — including remote locations, mobile users, and SaaS applications — through centralized management, consolidated core security tasks, and streamlined capabilities. For example, you should be able to use a single console to view all network traffic, manage configuration, push global policies, and generate reports on traffic patterns or security incidents. Your reporting capabilities must let your security personnel drill down into network, application, and user behavior for the context they need to make informed decisions.

When these capabilities are delivered from the cloud, your teams can build out the right security architecture to prevent known and unknown threats at every corner of your extended network environment. In today's constantly changing threat landscape, using a single security vendor to address the vast spectrum of your security and business needs isn't always practical. In this case, the ability to integrate with and consume third-party insight and innovation is critical. When evaluating future security vendors, be sure to evaluate the extensibility and programmability of what they offer.

Automate Routine Tasks and Focus on the Threats That Matter

There is a problematic shortage of cybersecurity skills within organizations worldwide. More than half of cybersecurity professionals recently surveyed by the Enterprise Strategy Group (ESG) confirm this opinion. This shortage is exacerbated by a dependency on too many manual processes for day-to-day security operations, such as chasing down data, investigating false positive alerts, and managing remediation. Manually analyzing

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

and correlating the vast number of security events slows mitigation, increases opportunities for error, and is impossible to scale. Security teams can easily drown in the volume of alerts and miss the critical, actionable ones. Although big data analytics uncovers hidden patterns, correlations, and other valuable insights to provide security teams with actionable intelligence, you still need the right data. That data must be sourced (and analytics-ready) from everywhere — networks, endpoints, SaaS applications, public clouds, private clouds, data centers, and so on.

TIP

Cybersecurity Ventures predicts that there will be 3.5 million cybersecurity job openings worldwide by 2021.

Using precise analytics to drive automation, you can easily implement security best practices like Zero Trust, streamline routine tasks, and focus on business priorities such as speeding application delivery, improving processes, and proactively hunting for threats in your environment. There are three capability areas that need to be automated on your firewall:

- **Workflow automation:** The firewall must expose standard application programming interfaces (APIs) so it can be programmed from other tools and scripts you may be using. In the cloud, it must integrate with tools like Ansible and Terraform. In addition, the firewall must be able to kick off workflows on other devices in your security ecosystem, using their APIs, without manual intervention.
- **Policy automation:** The firewall must be able to adapt policies to any changes in your environment, such as movement of applications across virtual machines. It must also be able to ingest threat intelligence from third-party sources and automatically act on that intelligence.
- **Security automation:** Your environment must be able to uncover unknown threats and deliver protections to the firewall so new threats are blocked automatically.

Some threats remain hidden in data. By looking deeper into that data across locations and deployment types, you can find threats that may be hiding in plain sight. With automation, you can accurately identify threats, enable rapid prevention, improve efficiency, better utilize the talent of your specialized staff, and improve your organization's security posture.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Consume New Security Innovations Easily

Consuming cybersecurity innovation is arduous. Organizations waste time deploying additional hardware or software every time they want to take advantage of a new security technology. They invest more resources managing their ever-expanding security infrastructure instead of improving their security controls to stay ahead of attackers and prevent threats.

As the number of needed security functions increases, there are two options: Add more siloed point security products or use an existing device to support new capabilities. If your firewall can act as a sensor and enforcement point for third-party technology, you can rapidly adopt new security innovations without deploying or managing endless new devices. Your firewall should enable teams to quickly discover, evaluate, and use new security technologies. Security teams should be able to collaborate between different applications, share threat context and intelligence, and drive automated response and enforcement with deeply integrated applications. In this way, they can solve the most challenging security use cases with the best technology available, and they can do so without the cost or operational burden of deploying new infrastructure for each new function.

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

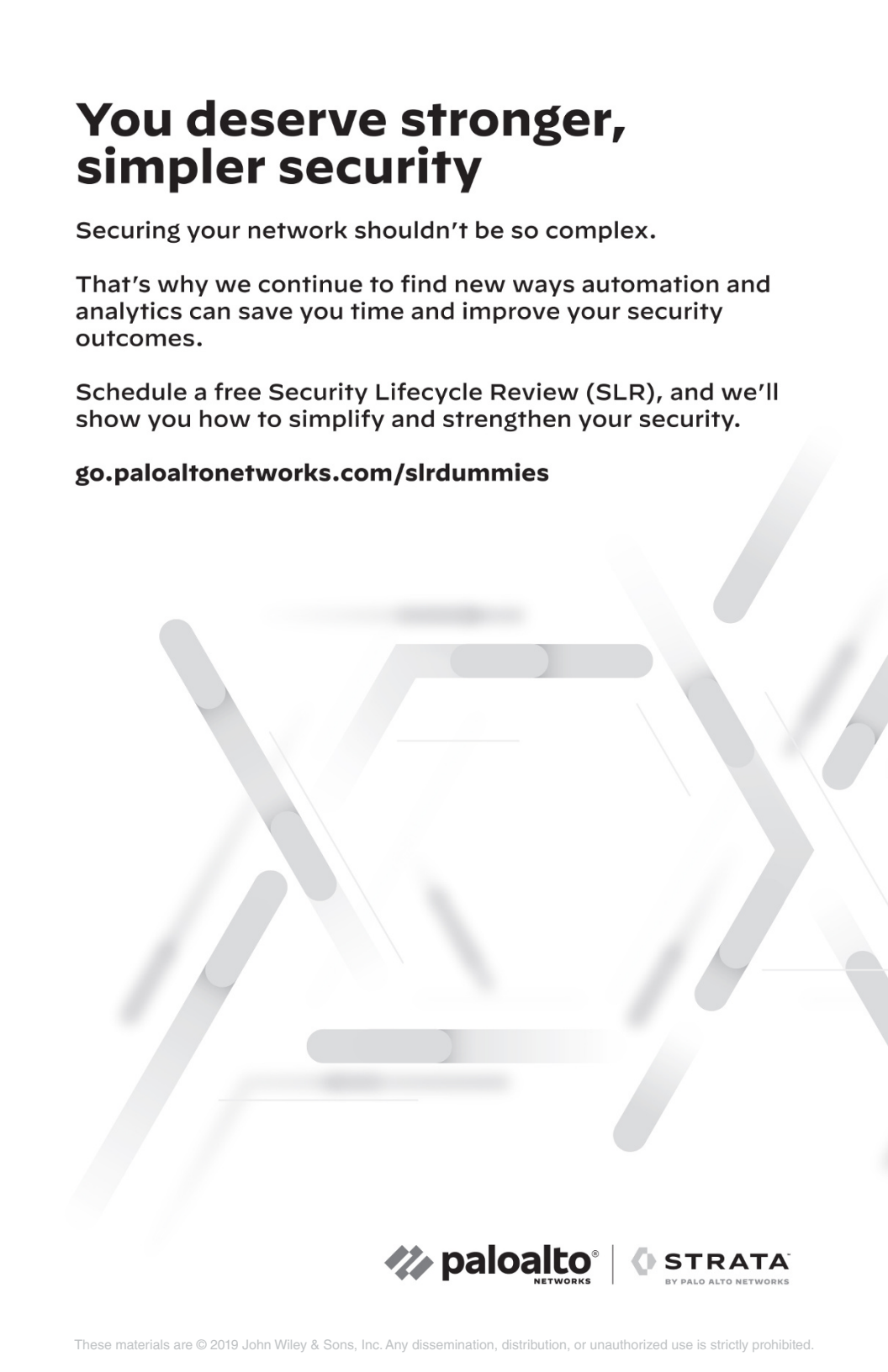
You deserve stronger, simpler security

Securing your network shouldn't be so complex.

That's why we continue to find new ways automation and analytics can save you time and improve your security outcomes.

Schedule a free Security Lifecycle Review (SLR), and we'll show you how to simplify and strengthen your security.

go.paloaltonetworks.com/slrdummies



paloalto® NETWORKS | STRATA BY PALO ALTO NETWORKS

These materials are © 2019 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Take back control of the apps, devices, and data on your network!

Traditional port-based firewalls are no longer adequate for network security. They were never designed to block evasive threats, inspect encrypted Internet traffic, or protect the cloud-based apps that are so common across corporate offices and branch locations, data centers, public and private clouds, and mobile devices. Security point products, such as intrusion prevention and web content filtering, provide siloed functionality, add complexity, and are no match for modern security threats and challenges.

Inside...

- Explore the evolving threat landscape
- Accelerate your journey to the cloud
- Simplify security management
- Implement a Zero Trust security approach
- Segment your network to reduce risk
- Learn what a next-gen firewall is – and isn't
- Discover next-gen firewall capabilities

paloalto NETWORKS

Lawrence Miller has worked in information technology in various industries for more than 25 years. He is the co-author of *CISSP For Dummies* and has written more than 150 *For Dummies* books on numerous technology and security topics.

Photo by chuttersnap on Unsplash

Go to Dummies.com®
for videos, step-by-step photos, how-to articles, or to shop!

ISBN: 978-1-119-62401-1
Not For Resale

for dummies® A Wiley Brand

Also available as an e-book

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.